2017

Molecular Evolution of Major Epidermal Structure Genes and an Integrative Transcriptome Analysis of Chicken Epidermal Embryogenesis

Weier Bao
University of South Carolina

MOLECULAR EVOLUTION OF MAJOR EPIDERMAL STRUCTURE GENES AND AN INTEGRATIVE TRANSCRIPTOME ANALYSIS OF CHICKEN EPIDERMAL EMBRYOGENESIS

by

Weier Bao

Bachelor of Medicine
Dalian Medical University, 2007

Master of Science
Kansas State University, 2010

Submitted in Partial Fulfillment of the Requirements

For the Degree of Doctor of Philosophy in

Biological Sciences

College of Arts and Sciences

University of South Carolina

2017

Accepted by:

Roger H. Sawyer, Major Professor

Lydia E. Matesic, Committee Member

Rekha C. Patel, Committee Member

Joseph M. Quattro, Committee Member

Wayne E. Carver, Committee Member

Cheryl L. Addy, Vice Provost and Dean of the Graduate School

DEDICATION

I dedicate my dissertation to my wife Flora and my lovely daughter Fiona.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my advisor, Dr. Roger Sawyer for his guidance during my PhD pursuit. He always encouraged me to explore my research areas and gave me his timely feedback. I am also thankful that he helped me to overcome my family hardship and gave me the best support. Special thanks go to Dr. Matthew Greenwold as my secondary mentor for his patient guidance and the discussions with him lead to my research inspirations. My research work would not have been possible without his help.

I would like to acknowledge Dr. Lydia Matesic, Dr. Rekha Patel, Dr. Joe Quattro and Dr. Wayne Carver for agreeing to serve on my committee members. I appreciate your patience shown in every committee meetings and thank you for your precious suggestions for my work. I also specially thank Dr. Carver's strong support and encouragement in my entire PhD period.

I would like to thank all the colleagues facilitating me to conduct my research. Many thanks to Dr. Richard Goodwin for providing chicken embryos, Dr. Diego Altomare for running the microarray experiment, Dr. Jeff Dudycha for providing the lab for RNA extraction work.

Finally, I want to say thank you to my family for your selfless support.

ABSTRACT

Alpha (α) and beta (β) keratins are the major structural proteins found in vertebrate epidermis and the β-keratins are only found in reptiles and birds. With the recent published 48 avian genomes, we searched and studied the molecular evolution of these gene families. We discovered that the expansion and contraction of different α- and β-keratins among the 48 phylogenetically diverse birds supports the importance of their role in the evolution of the feathers and the adaptation of birds to different ecological niches. Using a customized 44K microarray, we also performed transcriptome analysis on different epidermal regions (scutate scale, dorsal feather and wing feather) at important time points (day 8, 17 and 19) during chicken embryonic development. We profiled the differentially expressed α- and β-keratin genes in those comparison groups and demonstrated the important roles of α- and β-keratins in the development of the chicken epidermal appendages.

MicroRNAs have been found to widely regulate many biological processes in animals. Here we also utilized the 44K microarray transcriptome data to profile the miRNA expression during chicken embryonic development. With the application of various bioinformatic tools, based on the differentially expressed miRNA genes and mRNAs, we identified highly possible target genes for epidermal development in the chicken, and provided a rational for future miRNA target validation.

In previous studies, hundreds of genes (i.e., signaling pathway genes, structural genes, cell adhesion genes, etc.) have been associated with the morphogenesis of chicken epidermal structures as complex, interactive networks. A Weighted Gene Co-expression Network Analysis (WGCNA) using our microarray transcriptome data was performed to construct a co-expression network associated with traits. We identified two modules that were highly correlated with the developmental traits of the chicken scale and feather. The combination of traditional enrichment (KEGG and Gene Ontology) and novel enrichment (MSET and MeSH) analysis further demonstrated the important functional role of epidermal development related genes (EDRGs) and the hub genes to the development of scales and feathers. In the future, the discoveries of trait related modules will contribute to our understanding of the morphogenesis and differentiation of other epidermal appendages.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION AND BACKGROUND

A key event in support of the colonization of land by vertebrates was the evolution of a protective epidermis and its diversification into novel appendages (Sawyer et al., 1986; Alibardi and Sawyer, 2002; Greenwold and Sawyer, 2010; Strasser et al., 2014). These adaptations enabled successful interactions between the organisms and their new environments, such as thermoregulation, protection against excessive water loss and mechanical trauma as well as locomotion (Chuong and Homberger, 2003). The epidermis is a multi-layered epithelium, made up of keratinocytes, which are renewed by the proliferation of stem cells in the basal layer of the stratified epithelium. The evolution of the vertebrate epidermis and its appendages, such as claws, scales, beaks and feathers, was accompanied by molecular innovations of the structural components within the keratinocytes, as well as innovations of gene regulation via the developmental pathways during epidermal differentiation (Lowe et al., 2014). The genic and regulatory complement had been present in nonavian dinosaurs and contributed to the evolution of the emerging of pinnate feathers in dinosaurs (Lowe et al., 2014).

Alpha (α)-keratins are the major structural proteins found in the epidermis of all vertebrates, while beta (β)-keratins are only found in epidermis of reptiles and birds (Bell and Thathachari, 1963; Baden and Maderson, 1970; Haake et al., 1984; Rogers, 1985; Sawyer et al., 1986). The α-keratins are a type of intermediate filament, and form obligatory heterodimers of Type I (acidic) and Type II (basic) α-keratins (Berieter-Hahn et al., 1986; Greenwold et al., 2014). Beta (β)-keratins form a filament-matrix structure with the filament being 2-3nm in diameter, and containing a β-sheet-rich 34 amino acid residue domain, which is a common feature of all β-keratins (Sawyer et al., 2000; Fraser and Parry, 2008 and 2014; Hallahan et al., 2008; Dalla Valle et al., 2007, 2009 and 2010).

In birds, the β-keratins consist of four subfamilies (feather, claw, scale and keratinocyte), which are expressed in varying degrees in different epidermal appendages (Bereiter-Hahn et al., 1986; Rice et al., 2013; Smoak and Sawyer, 1983; Carver and Sawyer, 1988; Shames et al., 1989; Carver and Sawyer, 1989; Knapp et al., 1993). Both the α- and β-keratins have undergone gene duplication and functional diversification during vertebrate evolution (Vandebergh and Bossuyt, 2012; Greenwold and Sawyer, 2010; Greenwold and Sawyer, 2013; Greenwold et al., 2014; Li et al., 2013).

The sequencing of the chicken and zebra finch genomes (Hillier et al., 2004; Warren et al., 2010) demonstrated that the β-keratins make up a large gene family in these avian species. Numerous studies have characterized the β-keratin genes (Huth, 2008; Glenn et al., 2008; Greenwold and Sawyer, 2010; Greenwold and Sawyer, 2011; Greenwold and Sawyer, 2013; Greenwold et al., 2014). The comparison of the β-keratin genes in the chicken and zebra finch also demonstrated that they have similar genomic organization (Greenwold and Sawyer, 2010).

The expression of β-keratins in avian epidermal appendages, including scale, claw, feather, egg tooth, spur and beak, and their development have been characterized using electron microscopy, 2-dimentional protein gels and immuno-cytology (Sawyer and Abbott, 1972; Sawyer and Fallon, 1983; Shames et al., 1988; Lin et al., 2006). In scales, epidermal placodes begin to form by day 9 of incubation and by day 12 the scale ridge forms (Sawyer, 1972 a, b.; Sawyer and Abbott, 1972; Bereiter-Hahn et al., 1986). Expression of β-keratins in the developing scutate scale begins around day 14 of incubation in the embryonic epidermis and continues in the Beta Stratum of the adult scale (Shames and Sawyer, 1986). However, in the scaleless (sc/sc) mutant chicken,

which lacks scutate scales, β-keratin expression begins at 14 days of incubation, in the embryonic epidermis, but ceases by 17 days because the embryonic epidermis is lost at hatching (Shames and Sawyer, 1986). Since scales do not form in the mutant, a Beta Stratum, with its scale type β-keratins, does not develop (Shames and Sawyer, 1986).

Although chicken feathers start to develop as early as day 6 of incubation, the expression of β-keratin does not occur until around day 12.5 (Lucas and Stettenheim, 1972; Sengel, 1976; Haake et al., 1984). Beta (β)-keratin expression in feathers is continuous through hatching (Barnes, 1994; Shames et al., 1988). Presland et al., (1989a) found that at least five specific feather β-keratins are expressed in 14 day embryonic chickens. Glenn et. al., (2008) examined the evolutionary relationships among copies of feather beta keratin genes from several different orders of birds, and Greenwold and Sawyer (2010) used the newly sequenced genomes of the chicken and zebra finch to determine the genomic organization and molecular phylogenies of the beta keratin multi-gene family in these two birds.

Luo et al. (2012) compared early and late feathering in the day 1 post-hatched chicken by applying microarray technology. They found that at least 14 differentially expressed genes were significantly related to either α- or β-keratins. Therefore, in Chapter 2, I characterized the differentially expressed epidermal structural genes during the chicken embryonic development. We used a 44K chicken microarray, which was customized by Bao and Greenwold (Greenwold et al., 2014), to investigate the expression of α- and β-keratin genes during the embryonic development of scales and feathers in the chicken. With the availability of the newly sequenced genomes of 48 phylogenetically diverse birds (Zhang et al., 2014), in collaboration with Matthew Greenwold and others

involved with sequencing the genomes of 48 birds, we investigated the evolution of the avian epidermal structural genes (α- and β- keratins) by examining the gene copy number variation, the construction of molecular phylogenies, and the determination of the genomic orientation of the β-keratins (Greenwold, et al., 2014 – Bao is the co-first author).

MicroRNAs (miRNAs) have been drawing more attention from biologists since they were found to widely regulate biological processes (Betel et al., 2008). With their flexible targeting mechanism, I reasoned that the miRNAs may play a role in the development of the avian feather and scale. In the Chapter 3, I combined miRNA target prediction tools with high-throughput microarray expression data of miRNA genes and mRNA genes to identify miRNA-mRNA duplexes involved in the development of chicken epidermal appendages which have evolved to perform multiple important functions in birds. Furthermore, I have taken advantage of 60 genomes across Tetrapoda including the 48 bird genomes (Zhang et al. 2014) to investigate the evolution of chicken miRNA genes.

In chapter 4, I used a novel bioinformatic approach to examine gene networks in scale and feather development. Gene co-expression network analysis is a system biology approach to identify modules which cluster highly correlated genes (Langfelder et al., 2008). Weighted Gene Co-Expression Network Analysis (WGCNA) algorithm introduces topological overlap matrix to construct a more robust scale free network (Langfelder et al., 2008). In gene co-expression network, each gene corresponds to a node and the two genes are connected by an edge if their expression values are highly correlated. There are still large numbers of genes in the chicken that lack annotations or

have only limited annotations based on the prediction tools. By utilizing the customized 44K chicken microarray expression data, I was able to construct gene co-expression networks to identify gene modules highly correlated with chicken epidermal traits (feather or scale). In addition to the traditional Gene Ontology (GO) enrichment analysis, I applied the novel Modular Single-Set Enrichment Test (MSET) (Eisinger et al., 2013) and Medical Subject Headings (MeSH)-informed enrichment analysis (Lu, 2011; Morota et al., 2016) to demonstrate the validation of the WGCNA module construction by introducing external traits. Further enrichment analysis on the module hub genes with high connection with nodes also suggests the importance of biological role for the hub genes.

For many years, reptilian scales and avian scales have been considered to be homologous structures, as have the reptilian scales and avian feathers. In fact, studies have shown that the developmental signaling molecules (Wnt, Sonic Hedgehog) show similar patterns of expression during the early stages of scale development in the alligator and chicken, as well as in the early developmental stages of feathers (Harris et al. 2002; Musser et al., 2013). Most recently, it has been demonstrated that the anatomical "epidermal placode" characterizes the first step in the formation of reptilian scales, avian scale and feathers and the hair of mammals (Prum, 1999). Various scenarios have been suggested for the evolution of feathers from reptilian or avian scales (Prum, 1999; Prum and Brush, 2003; Sawyer and Knapp, 2003; Sawyer et al., 2005b). In the 1970s, it was first discovered that the scutate scales of the chicken developed from epidermal placodes similar to feathers (Sawyer, 1972 a, b). In addition, it was found that the scale epidermis of the chicken embryo produced and embryonic epidermis, which was discarded at

hatching (Sawyer et al., 1974 a, b). This embryonic epidermis is made up of independent cell populations originating from the basally located stem cell population. The first cell population generated is known as the periderm, which is found in all amniotes, and is characterized by unique peridermal granules that contain scaffoldin of the Epidermal Differentiation Complex (EDC) common to most vertebrates (Mlitz et al., 2014). The second cell population produced is known as the subperiderm, which has been extensively characterized in reptiles and birds (Sawyer and Knapp, 2003; Sawyer et. al, 2005b; Alibardi and Toni, 2008; Strasser et al., 2015). Using antibodies and in situ hybridization probes to the feather specific beta-keratins (Sawyer et al., 2005a), the Histidine-Rich Protein (HRP) of the EDC (Barnes and Sawyer, 1995; Alibardi et al., 2016), and Cysteine-rich Protein (EDCRP) of the EDC (Strasser et al., 2015), it was discovered that they are expressed in both the subperiderm of the embryonic epidermis and the barbules and barbs of the embryonic feather (Sawyer et al., 2003a; b; Sawyer et al., 2005a; Alibardi et al., 2006; Strasser et al., 2015; Alibardi et al., 2016). Unlike the independent subperidermal cell population of the apteric skin (general body skin) and epidermis of scales, which is lost at hatching, the subperidermal cells give rise to the barbules and barbs of the first downy feathers (Sawyer and Knapp, 2003; Sawyer et al., 2003a; Sawyer et al., 2005a; Strasser et al., 2015; Alibardi et al., 2016). These results strongly suggest that the evolution of the subperidermal population of cells, which originated in the ancestor of the archosaurians (crocodiles, dinosaurs and birds), gave rise to the barbs and barbules of the first feathers. This implies that the first feathers may have evolved directly from the epidermis without an intermediate scale step. In this study, the discovery of epidermal specific trait (scale or feather) related modules during chicken

embryogenesis provides the possible gene collaboration contributing to the distinct epidermal formation.

CHAPTER 2

DYNAMIC EVOLUTION OF THE ALPHA (α) AND BETA (β) KERATINS HAS ACCOMPANIED INTEGUMENT DIVERSIFICATION AND THE ADAPTATION OF BIRDS INTO NOVEL LIFESTYLES

Greenwold, M. J., Bao, W., Jarvis, E. D., Hu, H., Li, C., Gilbert, M. T. P., Zhang, G., Sawyer, R. H. (2014). *BMC Evolutionary Biology*, 14 (1), 249

ABSTRACT

Background

Vertebrate skin appendages are constructed of keratins produced by multigene families. Alpha (α) keratins are found in all vertebrates, while beta (β) keratins are found exclusively in reptiles and birds. We have studied the molecular evolution of these gene families in the genomes of 48 phylogenetically diverse birds and their expression in the scales and feathers of the chicken.

Results

We found that the total number of α-keratins is lower in birds than mammals and non-avian reptiles, yet two α-keratin genes (KRT42 and KRT75) have expanded in birds. The β-keratins, however, demonstrate a dynamic evolution associated with avian lifestyle. The avian specific feather β-keratins comprise a large majority of the total number of β-keratins, but independently derived lineages of aquatic and predatory birds have smaller proportions of feather β-keratin genes and larger proportions of keratinocyte β-keratin genes. Additionally, birds of prey have a larger proportion of claw β-keratins. Analysis of α- and β-keratin expression during development of chicken scales and feathers demonstrates that while α-keratins are expressed in these tissues, the number and magnitude of expressed β-keratin genes far exceeds that of α-keratins.

Conclusions

These results support the view that the number of α- and β-keratin genes expressed, the proportion of the β-keratin subfamily genes expressed and the diversification of the β-keratin genes have been important for the evolution of the feather and the adaptation of birds into multiple ecological niches.

BACKGROUND

The integument of amniotes has evolved from a basic cornified epidermis for protection against the environment and the retention of water into an elaborate covering with epidermal structures used additionally for sexual display, camouflage, locomotion, and thermoregulation (Gill, 1995). The claws, scales, beaks and feathers of reptiles and birds are formed from the products of two multigene families, alpha (α) and beta (β) keratins (Bell and Thathachari, 1963; Baden and Maderson, 1970; Haake et al., 1984; Rogers, 1985; Bereiter-Hahn et al., 1986). Alpha keratins, a subtype of intermediate filaments found in the epithelia of all vertebrates, have expanded and functionally diversified in amniotes through gene duplication (Vandebergh and Bossuyt, 2012). The β-keratins are found exclusively in reptiles and birds and have also expanded and diversified especially in the avian and chelonian lineages (Greenwold and Sawyer, 2010; Greenwold and Sawyer, 2013; Li et al., 2013).

The Type I (acidic) and Type II (basic/neutral) α-keratins form obligatory heterodimers (Lee and Baden, 1976; Hatzfeld and Franke, 1985) that make up the structural basis of the cornified epidermis and the epidermal appendages in mammals, such as wool, hair, claws, horns and hooves (Bereiter-Hahn et al., 1986; Vandebergh and Bossuyt, 2012; Fuchs and Marchuk, 1983; Rice et al., 2012). In birds, epidermal α-keratins make up the stratum corneum of the general epidermis and epidermal appendages such as the reticulate scale (O'Guin and Sawyer, 1982; Sawyer et al., 1986). They are present in varying degrees along with the β-keratins in the avian scutate scales, claws, beaks, spurs, and lingual nails (Bereiter-Hahn et al., 1986; Rice et al., 2012; Smoak and Sawyer, 1983; Carver and Sawyer, 1988; Shames et al., 1989; Carver and

Sawyer, 1989; Knapp et al., 1993). Although α-keratins are expressed in the early stages of feather development and in the cells of the rachis (Ng et al., 2012), the β-keratins make up 90% of the barbs and barbules of the mature feather (Haake et al., 1984; Walker and Rogers, 1976a and 1976b; Powell and Rogers, 1976; Alibardi and Sawyer, 2006; Alibardi, 2013; Kowata et al., 2014). In other words, the dynamic duplication and diversification of the β-keratin genes are thought to have contributed to the emergence of a novel epidermal appendage, the feather, which characterizes over 10,000 species of birds (Greenwold and Sawyer, 2010; Greenwold and Sawyer, 2013; Li et al., 2013).

The avian β-keratins were originally grouped into four subfamilies (claw, feather, feather-like, and scale β-keratins) based on expression profiles and sequence heterogeneity (Presland et al., 1989a and 1989b; Whitbread et al., 1991). More recently, an avian β-keratin isolated from cultured keratinocytes has been reported (Vanhoutteghem et al., 2004) and it is phylogenetically distinct from other β-keratin subfamilies (Greenwold and Sawyer, 2010; Greenwold and Sawyer, 2013; Vanhoutteghem et al., 2004; Dalla Valle et al., 2009a and 2009b; Greenwold and Sawyer, 2011). This keratinocyte β-keratin is also found in crocodilians, but not in the squamates examined to date (Greenwold and Sawyer, 2013). An additional β-keratin gene, BKJ, which is similar to feather-like β-keratins, has been identified on a unique locus and annotated as β-keratin from jun-transformed cells (Greenwold and Sawyer, 2010; Greenwold and Sawyer, 2013; Hartl and Bister, 1995). Thus, recent studies have regrouped the β-keratins into four different, but overlapping phylogenetically distinct subfamilies (claw, feather, scale, and keratinocyte β-keratins), proposing that the feather-

like and BKJ genes are basal genes within the feather β-keratin clade (Greenwold and Sawyer, 2013).

The Type I and II α-keratins are found on two unlinked genomic loci. In mammals, the Type I α-keratin locus is separated by a small cluster of keratin associated proteins (KAPs). However this Type I mammalian locus still shows a high level of synteny with the green anole lizard, chicken and zebra finch which lacks the KAPs (Vandebergh and Bossuyt, 2012). In birds, the Type I cluster is found on microchromosome 27 ((Vandebergh and Bossuyt, 2012; Hesse et al., 2004; Zimek and Weber, 2005). The Type II α-keratin cluster has been localized to linkage groups in the chicken and zebra finch genomes where they also show a high level of synteny with mammals and the green anole lizard (Vandebergh and Bossuyt, 2012). One Type I gene variant is found on the Type II cluster suggesting a common genomic locus of origin for the α-keratins in amniotes (Vandebergh and Bossuyt, 2012).

All four β-keratin subfamilies (claw, feather, scale, and keratinocyte β-keratins) have been localized to a single locus in both the chicken and zebra finch; microchromosome 25. However, several other unlinked loci contain feather β-keratins (Greenwold and Sawyer, 2010). Furthermore, the β-keratins from the green anole lizard are found on a single locus (Li et al., 2013; Greenwold and Sawyer, 2011; Alföldi et al., 2011) and nearly half of the western painted turtle β-keratins are found on a single locus that is syntenic to microchromosome 25 of the chicken and zebra finch suggesting a common ancestral locus for β-keratins (Li et al., 2013).

Here we have taken advantage of the sequencing of 48 bird genomes (Zhang et al., 2014) spanning the avian phylogeny see companion study, (Jarvis et al., 2014) to

investigate the evolutionary landscape of α- and β-keratins in the avian clade using copy number, molecular phylogenies, genomic orientation and transcriptome data. Our copy number data indicates that both α- and β-keratins have evolved in a dynamic manner with gene number contractions and expansions over the course of avian evolution leading to modern birds. Comparative transcriptome analyses demonstrate that 26 α-keratins and 102 β-keratins are differentially expressed in chicken scales and feathers during embryonic development. All four β-keratin subfamilies are highly expressed in developing scales, whereas the feather and keratinocyte β-keratins are highly expressed in the developing feather. The scales and feathers of birds have played important roles in the diversification of birds and their adaptation to multiple ecological niches. The dynamic evolution of the α- and β-keratins in the avian lineage accompanied these adaptations with the avian specific feather β-keratins making up to 85% of the total number of β-keratins, becoming the major structural component of the avian feather.

RESULTS

Genome searches of α-keratins show lineage specific gene losses and gains

We searched the genomes of 48 avian species that span the avian phylogeny, 2 crocodilians and 2 turtles (Zhang et al., 2014; Jarvis et al., 2014; St John et al., 2012; Wang et al., 2013) for α-keratins and made use of the α-keratin copy number estimates for the green anole, human, opossum, house mouse and platypus from Vandebergh and Bossuyt (2012) to test the hypothesis that there are no differences in copy number between birds and mammals and non-avian reptiles. We found that the total number of bird α-keratins, 26-38 ($\bar{x}$=31.271), is different from mammals and non-avian reptiles (fold change=0.707) with 29-62 ($\bar{x}$=44.222) α-keratins (Additional file 1; [Zhang et al., 2014]).

Furthermore, birds have a lower number of Type I α-keratins (range 10-18; $\bar{x}$=14.292) than mammals and non-avian reptiles (range 13-35; $\bar{x}$=23.22; fold change=0.616). The number of bird Type II α-keratins (range 11-22; $\bar{x}$=16.979) was also lower than that of mammals and non-avian reptiles (range 14-27; $\bar{x}$=21; fold change=0.809).

The differences in copy number of Type I and II α-keratins among vertebrate groups suggest a dynamic gain or loss of this gene family. To test this hypothesis, we annotated all of the bird, crocodilian, turtle, human and green anole α-keratins to determine which α-keratin genes may have been lost in the avian lineage (Additional file 2). We applied the same nomenclature of α-keratins as the one based on human/mammalian genes (Schweizer et al., 2006). We considered a gene to be lost in the avian lineage if it is present in human and at least one reptile and not found in any of the 48 bird genomes or avian expressed sequence tag (EST) libraries. Concurrently, we also identified the number of genes lost in the crocodilian and turtle lineages. Based on our annotations, we classified all avian and reptilian α-keratins into 17 Type I and 16 Type II genes (Table 2.1; tissue specificity in humans from Schweizer et al. (2006) and Moll et al. (2008). We found 8 Type I (see Table 2.1) and 6 Type II (see Table 2.1) α-keratins missing in birds. Interestingly, turtles have a relatively low number of Type I and II α-keratins and appear to have lost KRT10, KRT13, KRT31, KRT35 and KRT36 Type I genes and KRT1, KRT6A, KRT74, KRT79, KRT82 and KRT85 Type II genes. Crocodilians have also lost four of the same Type II genes, apparently independently (or possibly through incomplete lineage sorting of the common ancestor of turtles, birds and crocodiles) and lost the Type I gene KRT24. The cochleal or otokeratin Type II α-keratin cDNA was sequenced and described for the chicken by Heller et al. (1998) and appears to

have originated early in the reptilian lineage, as all reptiles and birds have at least one cochleal gene while none are found in humans.

Table 2.1. Type I and II α-keratin expression in humans

Tissue	Type I α-keratins	Type II α-keratins
Simple epithelium	**KRT10***	KRT8
	KRT18	
	KRT19*	
	KRT20	
Stratified epithelium	**KRT10***	KRT1
	KRT13	**KRT2**
	KRT14*	KRT5*
	KRT15*	KRT6A
	KRT16*	**KRT6B**
	KRT17*	
Corneal epithelium	KRT12	
Epithelium (nonspecific)	KRT23	KRT79
	KRT24	KRT80
	KRT42	
Hair	KRT14*	KRT5*
	KRT15*	**KRT72**
	KRT16*	KRT73
	KRT17*	**KRT74**
	KRT19*	KRT75
	KRT28	**KRT82**
	KRT31	KRT84
	KRT35	**KRT85**
	KRT36	

This table listed Type I and II α-keratin expression in humans and it is recreated from Schweizer et al. (2006) and Moll et al. (2008). Gene names in bold text have been lost in birds. *Indicates genes that are expressed in two of the tissues listed in the table.

While several α-keratin genes are absent in the avian lineage, at least one Type I and one Type II variant have expanded. KRT42, a Type I α-keratin found in epithelia of mammals has a mean copy number of 4.042 genes in birds and 1.143 genes in human and non-avian reptiles (fold change=3.54). KRT75, a Type II α-keratin gene associated with a feather rachis anomaly (Ng et al., 2012), is higher in birds ($\bar{x}$=8.333) relative to humans or non-avian reptiles ($\bar{x}$=4.714; fold change=1.76). At least two copies of KRT42 and KRT75 were found in all 48 birds.

The KRT75 expansion in birds led us to investigate whether an increase in copy number is related to mutations in KRT75 genes such as the mutation that causes the chicken frizzle feather phenotype (Ng et al., 2012). The associated mutation is at the exon 5/intron 5 junction that induces a cryptic splice site found within exon 5 of KRT75 resulting in a 69 base pair deletion. This cryptic splice site has a similar nucleotide motif (GTGAAG) as that of the normal splice site at the exon5/intron5 junction. A total of 286 α-keratins were annotated as KRT75 genes for the 48 bird genomes, with most species having 5 or more variants (Additional file 2). Of these 286, we found that only 12 KRT75 genes have the GTGAAG motif from 7 species (wild turkey, chicken, medium ground finch, zebra finch, American crow, chimney swift and domestic pigeon). Unexpectedly, these 7 species all had below the mean KRT75 copy number (8.333). These data suggest that the frizzle feather phenotype should be rare among birds. Indeed, only chickens, pigeons, geese and canaries have been described as having feather characteristics similar to the frizzle phenotype (Landauer and Dunn, 1930). This is consistent with our finding that the pigeon has the cryptic splice site as well as the two finches, which are in the same Family as canaries.

Avian adaptations to novel lifestyles was accompanied by β-keratin gene family dynamics

From all 48 bird genomes combined, we found 1623 complete β-keratin genes with both start and stop codons and without unknown sequence (-NNN-) or frame shift mutations. We also found 1084 incomplete avian β-keratins (Additional file 2). This analysis showed extreme variation in copy number for birds with the barn owl having only 6 β-keratins and the zebra finch having a maximum of 149 complete genes (Additional file 1;

[Zhang et al., 2014]). Consistent with earlier studies (Greenwold and Sawyer, 2013; Li et al., 2013), the American alligator and green sea turtle have 20 and 26 β-keratins, respectively. While, the barn owl and zebra finch represent the minimum and maximum number of total β-keratins, respectively, we found that the mean number of β-keratins in birds was 33.81. We identified 4 statistical outliers (zebra finch: 149, chicken: 133, pigeon: 81 and budgerigar: 71) that have a value greater than or equal to the third quartile plus 1.5 times the IQR (interquartile range), which are birds with the highest number of β-keratin genes (Additional file 1; [Zhang et al., 2014]). These drastic copy number differences may relate to the quality of the genome build or to other factors such as domestication, since they belong to 4 of 5 domesticated species among the 48 birds.

Annotation of the β-keratins was performed using the Greenwold and Sawyer (2010) dataset. The feather-like β-keratins and β-keratins from jun-transformed cells (BKJ) have been shown to group with feather β-keratins in previous phylogenetic analyses (Greenwold and Sawyer, 2013), therefore genes resulting in a best hit to those genes were annotated as feather β-keratins. We found that feather β-keratins comprised up to 85% of the total number of β-keratins for birds. Using Levene's test we can reject the null hypothesis of equal variance among claw ($W=23.442$, p-value<0.001), scale ($W=23.107$, p-value<0.001), keratinocyte ($W=21.744$, p-value<0.001) and total number of β-keratins, but not for feather β-keratins and the total number of β-keratins ($W=0.479$, p-value$=0.491$), indicating that the variance of feather β-keratin copy number for birds can be used as an indicator of the variance in the total β-keratin copy number for these 48 bird species.

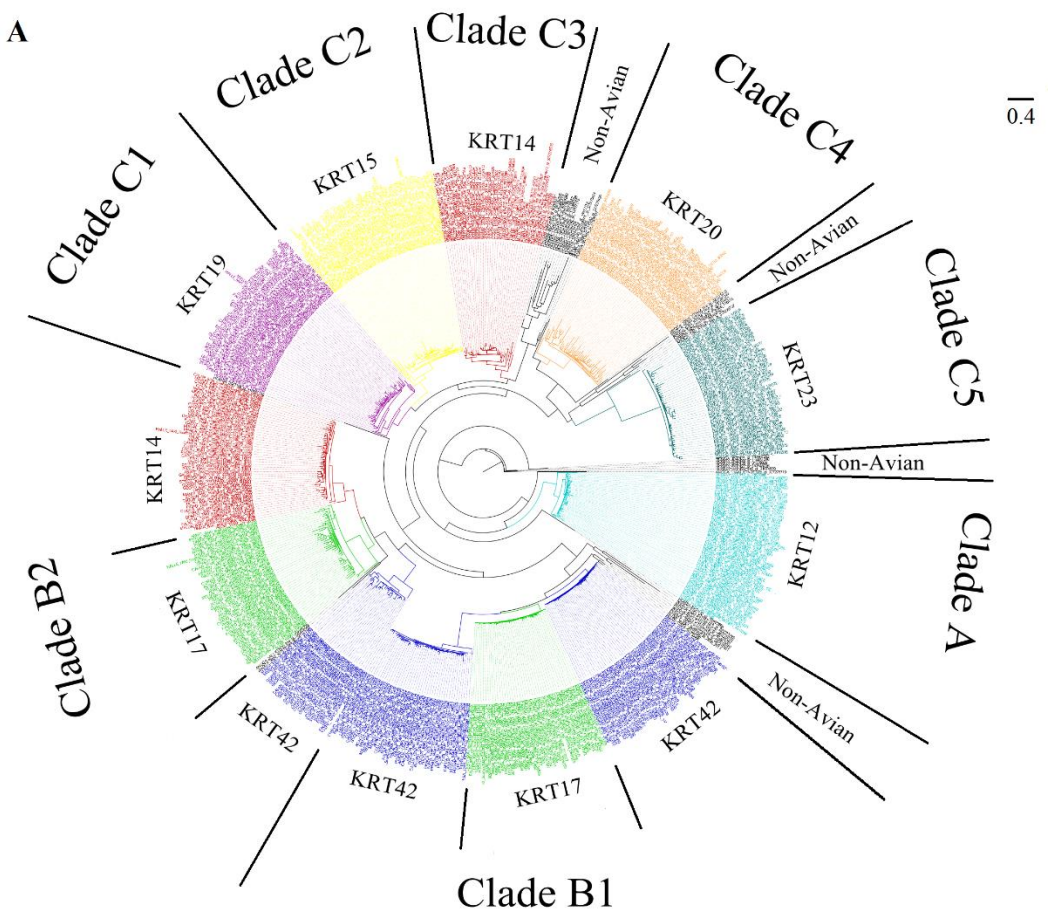
In order to ascertain if β-keratin copy number differences in birds correlate to species phylogenetic relatedness or lifestyle (aquatic and semi-aquatic and predatory; Additional file 1 [Zhang et al., 2014]; Additional file 3 [Jarvis et al. 2014]), we calculated the proportion of each of the four β-keratin subfamilies (Additional file 2) to the total number of β-keratins for each lifestyle using standard and phylogenetic ANOVA. We found that the proportion of feather β-keratins to the total number of β-keratins is significantly lower for aquatic and semi-aquatic birds (see Additional file 1; [Zhang et al., 2014]) than for land birds ($F_{1,46}=7.84$; standard ANOVA $p=0.007$; phylogenetic ANOVA $p=0.029$), while the proportion of keratinocyte β-keratins is significantly higher for aquatic and semi-aquatic birds than land birds ($F_{1,46}=10.79$; standard ANOVA $p=0.002$; phylogenetic ANOVA $p=0.013$). This includes aquatic and semi-aquatic species that are considered to have been independently derived according to the genome scale phylogeny in our companion study (Jarvis et al. (2014); Additional file 3) However, the species phylogeny (Additional file 3; [Jarvis et al., 2014]) indicates that the eagles do not group with the other aquatic birds, we therefore removed them from the aquatic bird list and found that only the higher proportion of keratinocyte β-keratins remained statistically significant ($F_{1,46}=6.84$; standard ANOVA $p=0.012$; phylogenetic ANOVA $p=0.043$) indicating more strongly that the change in β-keratin gene numbers could be associated with an aquatic lifestyle. We next considered birds with a predatory lifestyle (see Additional file 1; [Zhang et al., 2014]) and found that the proportion of claw β-keratins ($F_{1,46}=6.75$; standard ANOVA $p=0.0126$; phylogenetic ANOVA $p=0.033$) and keratinocyte β-keratins ($F_{1,46}=5.77$; standard ANOVA $p=0.02$; phylogenetic ANOVA $p=0.047$) is significantly higher in predatory birds than other birds while the proportion

of feather β-keratins ($F_{1,46}=7.81$; standard ANOVA $p=0.008$; phylogenetic ANOVA $p=0.022$) is significantly lower for predatory birds. Like the aquatic species, this finding occurs for independent lineages of predatory birds. We did not find any significant differences in copy number for the four β-keratin subfamilies between the major stem lineages of birds: Paleognathae vs. Neognathae or Paleognathae and Galloanserae vs. Neoaves; see also Jarvis et al. (2014). Together these data indicate that dynamic changes in the proportion of β-keratin subfamilies have occurred as birds have adapted to novel lifestyles.

The α- and β-keratin multigene families have similar patterns of sequence divergence following gene duplication

To further our understanding of the evolution of the α- and β-keratin multigene families in birds we performed phylogenetic analyses and examined their genomic orientation (Figure 2.1- 2.3, Additional data file 4; [Zhang et al., 2014]). By examining these two types of data we were able to elucidate the gene duplication history of multigene families and gain a deeper understanding of their genomic origin.

The phylogenetic analyses of the Type I and II α-keratins (Figure 2.1A and 2.2A) demonstrate that they can be separated into 3 main clades (Clade A, B and C) with Clade A being composed of a single basal gene. The remaining two clades (Clade B and C) of the Type I and Type II α-keratins are composed of multiple phylogenetically significant sub-clades of different gene variants. The α-keratins genes, KRT75 and 42 are distributed among several sub-clades in their respective phylogenies (Figure 2.1A and 2.2A) and the



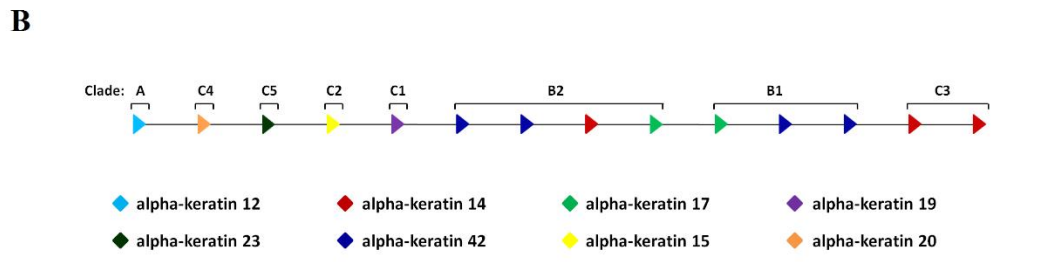
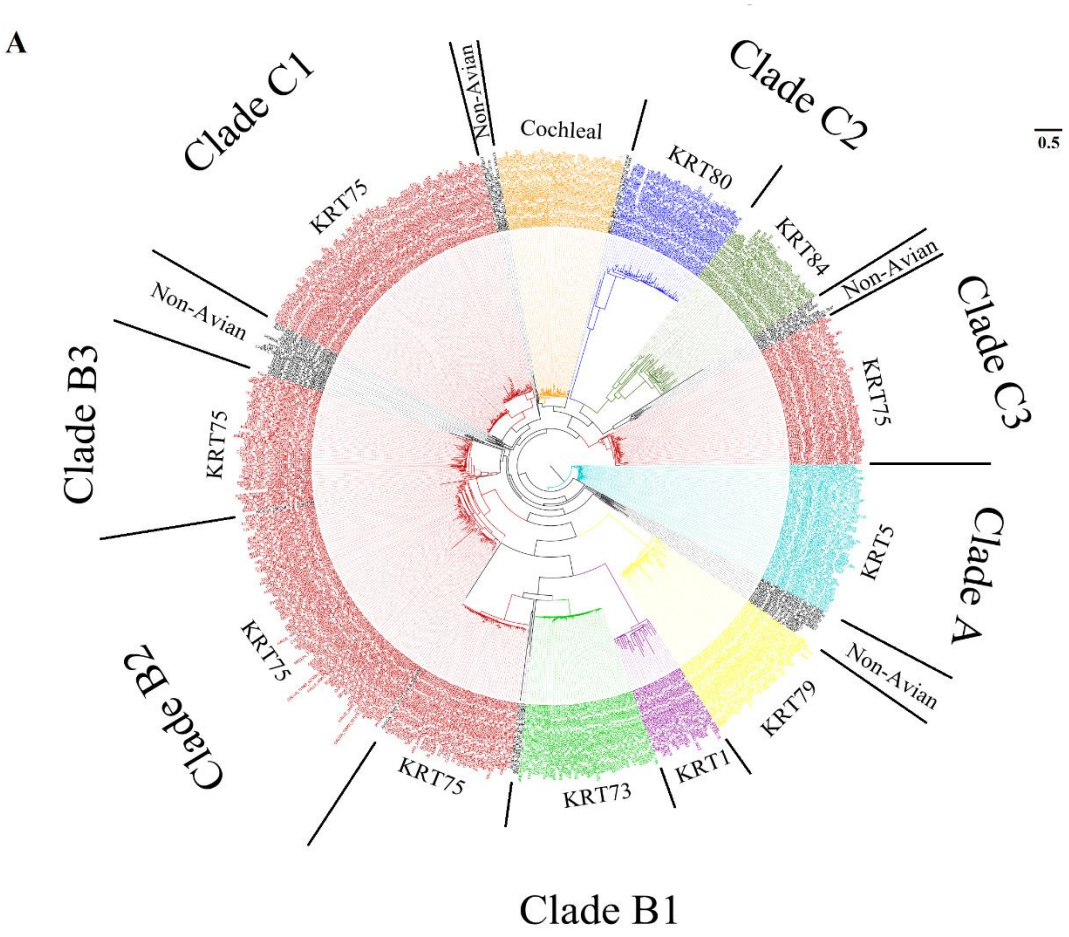


Figure 2.1. Molecular phylogeny and proposed genomic orientation of Type I α-keratins. Part A is the maximum likelihood phylogeny of Type I α-keratins from human, green anole lizard, green sea turtle, American alligator and the 48 birds. Annotation of Type I α-keratins is based upon avian gene annotations. All clades are statistically significant. Genes labeled as non-avian include genes from human, green anole lizard, green sea turtle and/or American alligator. Part B is the proposed genomic orientation of Type I α-

keratins in birds. While this whole region was not found on a single continuous genomic scaffold for some birds, the genomic alignment of scaffolds/contigs with at least 2 different gene variants resulted in this proposed consensus gene orientation of birds. Annotations are based on Part A. The direction of the arrow is indicative of the DNA strand.



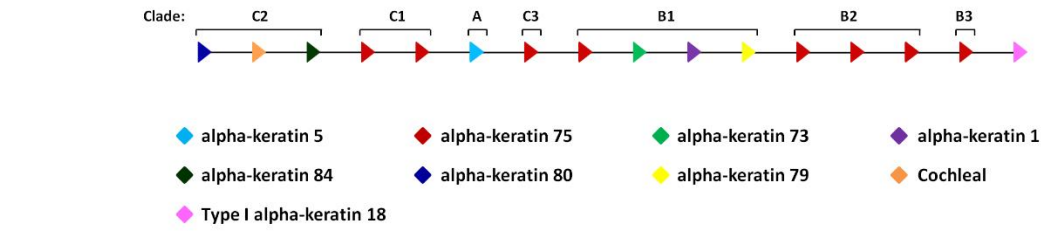


Figure 2.2. Molecular phylogeny and proposed genomic orientation of Type II α-keratins. Part A is the maximum likelihood phylogeny of Type II α-keratins from human, green anole lizard, green sea turtle, American alligator and the 48 birds. Annotation of Type II α-keratins is based upon avian gene annotations. All clades are statistically significant. Genes labeled as non-avian include genes from human, green anole lizard, green sea turtle and/or American alligator. Part B is the proposed genomic orientation of Type II α-keratins in birds. While this whole region was not found on a single continuous genomic scaffold for some birds, the genomic alignment of scaffolds/contigs with at least 2 different gene variants resulted in this proposed consensus gene orientation of birds. Annotations are based upon Part A. The direction of the arrow is indicative of the DNA strand. The one Type I α-keratin, KRT18, shown in the consensus genomic orientation was only found in 7 species of birds, but was included in this figure based on the present data and previous studies (Vandebergh and Bossuyt, 2012; Hesse et al., 2004; Zimek and Weber, 2005).

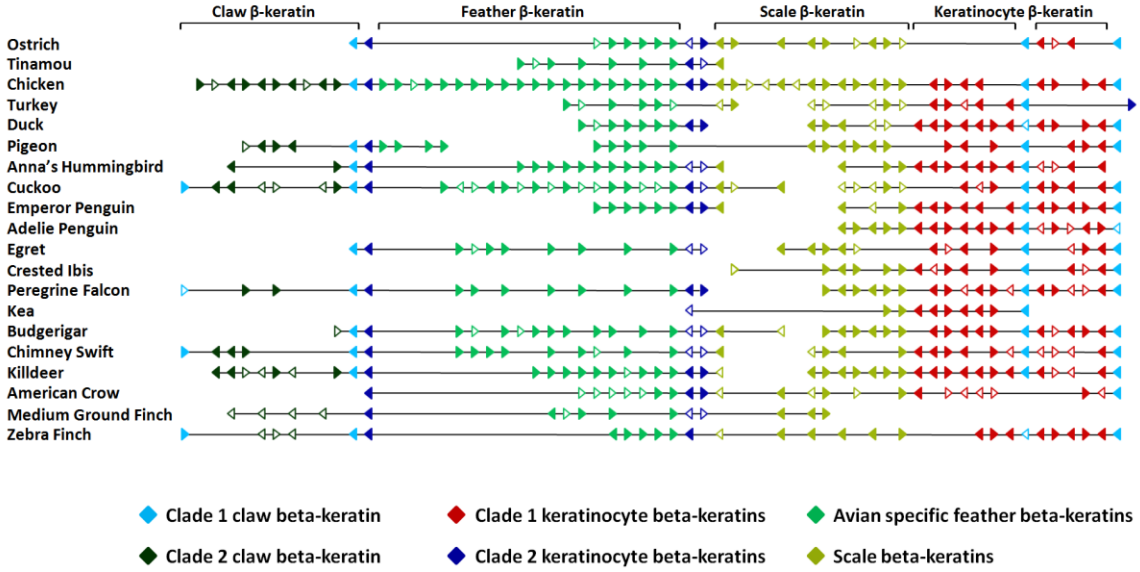


Figure 2.3. Genomic orientation of β-keratins in birds. This figure is a genomic alignment of β-keratins in birds containing a genomic locus with at least two β-keratin subfamilies. For the chicken and zebra finch this locus is microchromosome 25. Although feather β-keratins can be found on many genomic loci other than the one shown here (Greenwold and Sawyer, 2010), we focused on this locus as it has members from all of the β-keratin clades. The annotations are based on our β-keratin phylogeny (Additional file 4; [Zhang et al., 2014]). The breaks in the line for each species are indicative of different genomic scaffolds. The direction of the arrow is indicative of the DNA strand. The arrows with solid colors are complete genes and those with white centers are incomplete genes. The annotation above the figure is based on each of the four β-keratin subfamilies, while the clades based on the β-keratin phylogeny (Additional file 4; [Zhang et al., 2014]) are shown below the figure.

genes in different sub-clades are generally found interposed between other α-keratin genes on their respective α-keratin loci (Figure 2.1B and 2.2B). These phylogenies clearly demonstrate that KRT42 and KRT75 genes have expanded and their duplication history is marked by non-tandem duplication and subsequent sequence divergence. Our phylogenetic and genomic orientation data further support the idea that Type I and II α-keratins have evolved through gene duplications in a concerted fashion (Blumenberg, 1988).

For the β-keratin phylogeny, the green anole forms a clade and therefore was selected as the outgroup (Additional file 4; [Zhang et al., 2014]). Furthermore, a clade of keratinocyte β-keratins composed of bird, turtle and crocodilian genes forms the basal sister clade (Clade 1 keratinocyte β-keratins, Additional file 4; [Zhang et al., 2014]). The remaining bird, turtle and crocodilian β-keratins form a second major clade that is composed of another keratinocyte β-keratin clade (Clade 2 Keratinocyte β-keratins), two claw β-keratin clades, one scale β-keratin clade and an avian specific feather β-keratin clade. While all of the scale β-keratins annotated in Additional file 4 [Zhang et al., 2014] group together, only a portion of them form a phylogenetically significant clade with strong bootstrap support. Similar to the α-keratin genes, KRT42 and KRT75, the β-keratin subfamilies, keratinocyte and claw β-keratins, have duplicated in a non-tandem pattern (Figure 2.3) and form multiple phylogenetically significant clades (Additional file 4; [Zhang et al., 2014]). Collectively, these results indicate that the transposition of duplicated genes on the same locus in a non-tandem fashion is adequate to induce a relatively high level of sequence divergence possibly resulting in neofunctionalization.

Generally our gene trees (Figure 2.1A, 2.2A and Additional data file 4; [Zhang et al., 2014]) did not follow the genome scale phylogeny of the 48 bird species in Jarvis et al. (2014) (Additional file 3). This is not surprising because in that study no single gene tree they analyzed was identical to the species tree. Due to incomplete lineage sorting, most genes differed from the species tree by 20% of the branches, and our findings above indicate large scale convergence of keratins among aquatic and predatory birds with relationships different from the species tree. For each gene phylogeny (Figure 2.1A, 2.2A and Additional data file 4; [Zhang et al., 2014]), however, we found that sub-clades of closely related species frequently grouped together such as those of the Palaeognathae (tinamou and ostrich), Galloanseres (chicken, turkey and duck) and some species of the Psittacopasserimorphae (songbirds and parrots).

The phylogenetic analyses of Type I and II α-keratins (Figure 2.1A and 2.2A) support the interpretation that birds have lost 14 α-keratin gene variants (see Table 2.1). We found that the gene variants that are missing in birds form statistically significant clades in human and non-avian reptiles. Additionally we found that our phylogenetic analyses resulted in slightly different copy number counts for the α-keratin genes and β-keratin subfamilies from the annotations detailed above (see Additional file 2). However, the statistical significance of the comparisons between the β-keratin subfamilies and lifestyles largely remained valid with the phylogeny data (data not shown). The exception is with the keratinocyte β-keratin comparison between aquatic and land birds ($F_{1,46}=5.76$; standard ANOVA $p=0.021$; phylogenetic ANOVA $p=0.057$) when the eagles are grouped with land birds.

Previous studies (Vandebergh and Bossuyt, 2012; Hesse et al., 2004; Zimek and Weber, 2005) have shown that one Type I α-keratin gene (KRT18) is found on the Type II α-keratin locus for fish, amphibians, mammals and the green anole lizard indicating that the α-keratins evolved from a single locus. While Vandebergh and Bossuyt (2012) found a Type I α-keratin on chromosome unknown of the chicken, they found no direct evidence that Type I and II α-keratins are linked in birds. In fact the gene they found on chicken chromosome unknown was not found during our genome searches. However, we did find that 8 of the bird species (chimney swift, common cuckoo, little egret, peregrine falcon, crested ibis, brown mesite, white-throated tinamou and common ostrich), the 2 crocodilian species and the western painted turtle had one Type I α-keratin on a Type II locus (Figure 2B). Furthermore, these genes were annotated as KRT18. The green sea turtle and green anole lizard both have 2 KRT18 genes, which are found on 2 different loci.

Differential expression of the α- and β-keratin genes in chicken epidermal tissue during embryogenesis

We performed transcriptome analyses of scale, dorsal feather and wing feather tissues during chicken development using a customized version of the chicken 44K Agilent microarray (Li et al., 2008). We customized the microarray chip by adding all 27 α-keratins and 102 of the 133 chicken β-keratins. The number of unique 60-mer oligonucleotides of β-keratins was constrained due to the highly repetitive nature of feather β-keratins and thus we were only able to produce unique oligonucleotides for 68 of the 99 chicken feather β-keratins.

Tissue samples from the chicken scutate scale, dorsal feather and wing feather were from embryonic day 17 and 19 and scutate scale and dorsal feather at day 8. Although feather morphogenesis begins as early as day 6.5 to 7, the cellular differentiation of barbs and barbules and the accumulation of β-keratin does not begin until ~ day 12 of embryogenesis (Bell and Thathachari, 1963; Haake et al., 1984; Lucas and Stettenheim, 1972; Sengel, 1976). Scutate scale morphogenesis does not begin until day 9.5 of embryogenesis, and β-keratin accumulation is not detected until 15-16 days of development (Sawyer et al., 1986; Sawyer, 1972; Sawyer and Fallon, 1983; Shames and Sawyer, 1986; Shames and Sawyer, 1987). Thus, we selected day 8 for the initial sampling of the scale and feather tissues (Sawyer, 1972).

Comparison of day 8 scutate scale and day 8 dorsal feather tissues showed that 6 α-keratins and 18 β-keratins were differentially expressed, but the fold change values were 5 or below (Table 2.2 and Additional file 5). Comparisons of day 8 and day 17 had the largest number of differentially expressed α- and β-keratins. In comparing day 8 dorsal feather to day 17 dorsal feather we found 20 up-regulated and 2 down-regulated α-keratins in the day 17 dorsal feather. Also 98 β-keratins were up-regulated in day 17 dorsal feather, which showed up to an 112,000 fold change for a keratinocyte β-keratin and a 5,000 fold change for a feather β-keratin. We found 23 α-keratins that were up-regulated and 2 α-keratins that were down-regulated in day 17 scutate scale while 101 β¬-keratins were up-regulated in day 17 scale (Table 2.2). Day 8 and 19 comparisons of the dorsal feather and scutate scale had slightly lower numbers of α- and β-keratin indicating that day 17 of development may be the peak level of keratin expression.

Table 2.2. Expression of α- and β-keratins during embryonic chicken development.

Sample comparisons (Tissue/embryonic day of development)	Number of differentially expressed α- and β-keratins					
	Type I α-keratins*	Type II α-keratins*	Keratinocyte β-keratins*	Scale β-keratins*	Claw β-keratins*	Feather β-keratins*
DF8/DF17	11[↓]	2[↑];7[↓]	10[↓]	9[↓]	12[↓]	67[↓]
DF8/DF19	11[↓]	1[↑];8[↓]	9[↓]	5[↓]	8[↓]	61[↓]
DF17/DF19	2[↑]	0	3[↑]	8[↑]	10[↑]	67[↑]
DF17/WF17	5[↓]	1[↓]	2[↓]	7[↓]	0	3[↓]
DF19/WF19	1[↑];1[↓]	3[↑];1[↓]	0	8[↓]	8[↓]	5[↓]
SC8/DF8	1[↑];1[↓]	3[↑];1[↓]	0	2[↑]	0	13[↑];1[↓]
SC8/SC17	1[↑];12[↓]	1[↑];10[↓]	10[↓]	9[↓]	12[↓]	70[↓]
SC8/SC19	1[↑];13[↓]	1[↑];8[↓]	10[↓]	9[↓]	12[↓]	65[↓]
SC17/DF17	4[↑];2[↓]	6[↑]	6[↑]; 1[↓]	9[↑]	11[↑]	17[↑]; 8[↓]
SC17/SC19	1[↑];1[↓]	2[↑]	4[↑]	3[↑]	1[↑]	35[↑]
SC17/WF17	4[↑];1[↓]	6[↑];1[↓]	9[↑]; 1[↓]	4[↑]; 4[↓]	12[↑]	17[↑]; 21[↓]
SC19/DF19	7[↑];1[↓]	5[↑]	7[↑]; 2[↓]	9[↑]	11[↑]	26[↑]
SC19/WF19	6[↑];1[↓]	4[↑]	4[↑]; 3[↓]	4[↑]	10[↑]	3[↑]; 1[↓]
WF17/WF19	1[↑]	0	1[↑]	0	0	5[↑]

This table listed the expression of α- and β-keratins during embryonic chicken development. Number of differentially expressed α- and β-keratins during embryonic chicken development for 14 sample comparisons with a p-value cutoff of 0.05 and a fold change cutoff of 2.0. DF, dorsal feather; WF, wing feather; SC, scale. *Direction of selection is indicated in brackets (↑: Up-regulated; ↓: Down-regulated) after each copy number and refers to the first sample for each comparison.

Although the scale β-keratins were annotated based upon their expression in scale tissue (Presland et al., 1989b), it appears that the claw β-keratins are expressed at the highest level in scale tissue. In the scutate scale comparisons of day 8 vs. 17 and day 8 vs. 19, 7 out of the 10 highest fold changes (up-regulated genes) in day 17 and 19 scutate scale are claw β-keratins. Additionally, the day 17 and 19 scutate scale inter-tissue comparisons (dorsal and wing feather) showed that 9 of the highest fold changes (up-regulated genes) in the scutate scale are claw β-keratins indicating that claw β-keratins have an important role in the composition of epidermal appendages, such as scales, in addition to the claw and beak (Greenwold and Sawyer, 2010).

Four sample comparisons had genes from the feather β-keratin subfamily that were up and down regulated (Table 2.2). The day 8 comparison of the scutate scale and dorsal feather indicates that a single feather β-keratin from microchromosome 27 is up-regulated in the day 8 dorsal feather while all of the down-regulated feather β-keratins are on different loci (chromosome 1, 2, microchromosome 25, and chromosome unknown). Furthermore, feather β-keratins on microchromosome 27 are up-regulated in day 17 scutate scale in comparisons of the scutate scale day 17 and dorsal and wing feather day 17. Additionally, comparisons between the scutate scale and wing feather during day 19 of embryogenesis show an up-regulation of microchromosome 27 feather β-keratins and down-regulation of feather β-keratins on other loci in the day 19 scutate scale tissue. These data indicate that feather β-keratins on microchromosome 27 are regulated differently from feather β-keratins on other loci (chromosome 1, 2, 6, 10, and unknown and microchromosome 25).

We found that the basal BKJ genes of the feather β-keratin clade (Additional file 4; [Zhang et al., 2014]) are expressed at a higher level in the dorsal and wing feather when compared to the scutate scale at day 17 and 19 (Additional file 5). Although BKJ genes are expressed in higher levels in the feather, they are also expressed in the scutate scales as evidenced by the down-regulation in the scutate scale comparisons between day 8 and day 17 and 19. The feather-like β-keratins are found in multiple comparisons indicating they are expressed in both feather and scutate scale tissue. Interestingly, the only three feather β-keratins expressed in dorsal and wing feather comparisons at day 17 are the feather-like β-keratins suggesting that they have an important role in feather morphology. While the feather-like β-keratins are linked to other feather β-keratin genes

on chromosome 25, the BKJ genes are found on chromosome 6 and are not linked to any other β-keratins indicating that intra and inter-locus differential expression occurs among the feather β-keratin clade.

Feather β-keratins in the chicken genome are found on multiple loci (Additional file 4; [Zhang et al., 2014]) (Greenwold and Sawyer, 2010). Based on our sample comparisons in this study we were able to determine which feather β-keratins from which chicken genomic loci (GALGA, chromosomes) were being expressed in the dorsal and wing feathers during embryonic development. The genomic loci of feather β-keratins being expressed in day 17 dorsal (down) feathers and day 17 wing feathers are summarized in Figure 2.4. The feather β-keratins expressed in the day 17 down feathers are located on GALGA 1, 6, 10, 25 and 27. In addition, the feather-like genes on GALGA 25 were expressed as was feather β-keratin on GALGA unknown. The feather β-keratins expressed in the day 17 wing feathers are located on GALGA 6, 10, and 25. In addition, the feather-like genes on GALGA 25 were expressed as was feather β-keratin on GALGA unknown (Figure 2.4).

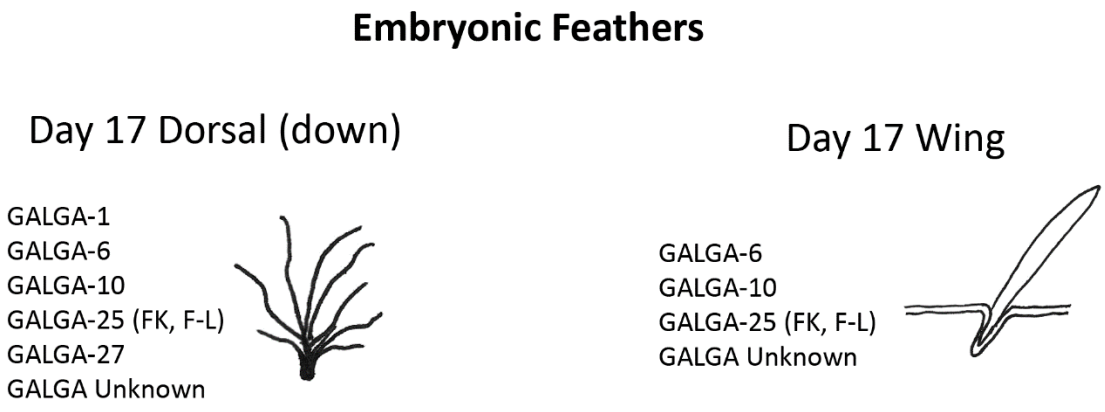


Figure 2.4. Expression of feather β-keratins during embryonic feather development. This figure summarizes the present data on the expression of feather β-keratin by chromosomal location in embryonic (day 17 down and wing feathers) feathers using data from the present study. For each feather type (Day 17 down, Day 17 Wing), the chicken (GALGA) chromosome number of the feather β-keratins expressed is listed. The feathers of day 17 dorsal skin express feather β-keratins located on GALGA chromosome 1, 6, 10, 25 (both feather and feather-like β-keratins) and chromosome unknown. Wing feathers, at day 17, express feather β-keratins from GALGA chromosome 6, 10, 25 (both feather and feather-like β-keratins) and chromosome unknown.

The only comparison showing both up and down-regulated scale β-keratins is the day 17 scutate scale versus wing feather, which indicates that scale genes annotated as 1, 2, 3 and 5 are up-regulated while scale 7, 8, 9 and 10 are down-regulated. These scale β-keratins are all found on the same locus (GALGA 25) and their number describes their orientation in a 5' to 3' direction. Alternatively, only one keratinocyte β-keratin (GALGA25_Ktn6) is consistently differentially expressed from the other keratinocyte β-keratins on GALGA 25. These results indicate that while β-keratins from all subfamilies are being expressed in these tissues, intra-locus differential expression of GALGA 25 β-keratins and inter-locus differential expression of feather β-keratins may contribute to the structural complexity of these and other avian epidermal appendages.

DISCUSSION

This study made use of the newly published genomes of 45 birds in addition to the 3 previously published bird genomes (chicken, zebra finch and turkey) to investigate the multigene families of α- and β-keratins (Zhang et al., 2014). Incomplete or low coverage genomes can lead to an underestimate of gene family copy numbers. For α-keratins, little variation is seen among our copy number estimates (Additional file 1; [Zhang et al., 2014]). Furthermore, based on the α-keratin annotation, we find consistent copy number estimates of the different Type I and II α-keratins (Additional file 2). In contrast to the α-

keratins, the β-keratins have a much larger variation in copy number estimates (Additional file 1; [Zhang et al., 2014]). While both of these gene families are tandemly arrayed on at least 2 genomic loci strong differences exist in the variation of the copy number estimates among the 48 birds. The newly sequenced bird genomes are separated into two coverage groups; low (<50X coverage) and high (>50X coverage) [Zhang et al., 2014]. The coverage in these two groups vary, but if the β-keratin copy number is related to genome coverage the copy number of β-keratins should correlate with fold coverage and contig and scaffold N50. However, for each fold coverage group (low and high) we do not find a statistically significant correlation between β-keratin copy number and fold coverage, contig N50 or scaffold N50. While, we do not discount the likelihood that some of the bird species in this study have unsequenced β-keratins, we believe that the relative variation in β-keratin copy number among birds is appropriately represented in this study.

Alpha (α)-keratins

The α-keratin nomenclature used in this study is based upon mammals and more specifically humans (Schweizer et al., 2006). While mammals have shown the largest expansion of α-keratins among amniotes (Vandebergh and Bossuyt, 2012), we find that there is avian specific gene loss and gain of α-keratins. The expansion of specific α-keratin gene variants (KRT42 and KRT75) in birds may not be the result of gene duplication of a single "parent" gene, but instead the duplication of several different gene variants resulting in novel α-keratins of avian origin. If some of these genes are novel α-keratins as the phylogeny and genomic alignment data indicate, then the current α-keratin nomenclature, based on mammalian genes, does not adequately account for the diversity

found in birds. Therefore, we suggest that the KRT42 genes be annotated as KRT42a and b and the KRT75 genes be annotated as KRT75a-e to reflect their phylogenetic relationship and genomic orientation.

Our discovery of the KRT42 and KRT75 expansion in the avian lineage and Ng et al. (2012) discovery that KRT75 is important in feather rachis development indicates that the duplication of KRT42 and KRT75 α-keratins may be the result of concerted evolution and that together they form the α-keratin heterodimer in feathers. Furthermore, it is likely that these duplicated genes contributed to the evolution of feathers as did the feather β-keratins (Ng et al., 2014).

Beta (β)-Keratins

The fact that the extreme statistical outliers from the average number of ~34 β-keratins across bird species are all species that have undergone various degrees of domestication (zebra finch, chicken, pigeon and budgerigar), indicate that there could be an association between these observations. In support of this relationship, both the Peking duck (46) and turkey (46), the remaining domesticated species among the birds, have an above average number of β-keratins. Given that domestication may increase recombination rate (Burt and Bell, 1987; Ross-Ibarra, 2004) the extreme variation in β-keratin copy numbers among birds may be partially linked to higher recombination rates on β-keratin loci and the domestication of these species. The differential expression of feather β-keratins is related to their genomic locus (Ng et al., 2014), signifying that expansion of feather β-keratins, through unequal crossing over events on specific loci, may be induced by artificial selection.

Expression of feather β-keratins and the evolution of feathers

Numerous studies have examined the biochemical and molecular make up of embryonic and adult feathers, as well as their component parts (Sawyer et al., 1986; Walker and Rogers, 1976a and 1976b; Alibardi and Sawyer, 2006; Alibardi, 2013; Kowata et al., 2014; Dalla Valle et al., 2009a and 2009b; Dalla Valle et al., 2010;Shames and Sawyer, 1987; Ng et al., 2014;Kemp and Rogers, 1972; Walker and Bridgen, 1976; Kemp, 1975; Gregg and Rogers, 1986; Shames et al., 1988). For example Kemp (1975) suggested that there were 25-35 different feather keratin mRNA molecules in the embryonic feather and a total of 100-240 keratin genes in the chicken genome. The present study supports the view that a high number of α- and β-keratins are expressed during the embryonic development of scutate scales and feathers in the chicken (Table 2.2). These results are further supported by a recent study by Ng et al. (2014), which found that 90% of α-keratin and over 95% of β-keratin genes in the chicken are differentially expressed during post-hatching feather genesis. However, the number of β-keratins that can be extracted from the cornified tissues of scales and feathers and detected on 2-dimensional gels is considerably smaller (Rice et al., 2012; Knapp et al., 1993; Shames and Sawyer, 1986; Gregg and Rogers, 1986; Shames et al., 1988; Knapp et al., 1991), suggesting that messenger RNAs are being inactivated, perhaps by microRNAs (Zhang et al., 2013).

Recently, Kowata et al. (2014) found that the feather β-keratin on chromosome 7 of the chicken (GALGA 7) is expressed in the cells that form the barbules of pennaceous feathers but is not expressed in the barbules of plumulaceous feathers. In the present study, we did not find any differential expression of the GALGA 7 feather β-keratin in embryonic feathers supporting the results of Kowata et al. (2014). However, we did find that GALGA 27 feather β-keratins are differentially expressed in comparison to feather

β-keratins on other loci and that GALGA 27 feather β-keratins are generally up-regulated in scale tissue. Previous studies have demonstrated that the ancestral locus of β-keratins is homologous to GALGA 25 of the chicken (Greenwold and Sawyer, 2010; Greenwold and Sawyer, 2013; Li et al., 2013), suggesting that feather β-keratins diversified to other genomic loci through duplication and translocation. Recently, Ng et al. (2014) examined which genomic loci (chromosome) are utilized for the expression of feather β-keratins in post-hatched contour and flight feathers. While it is clear that feather β-keratin located on GALGA 7 is only expressed in the barbules and possibly the hooklets of pennaceous feathers, feather β-keratins from multiple loci are expressed in the ramus, rachis, and calamus of post-hatched feathers. Overall these data suggest that as the avian epidermis evolved to produce novel structures (such as pennaceous feathers) it took advantage of the diversity of feather β-keratins that evolved on different loci.

Evolution of birds into novel ecological niches

Bird diversification is marked by evolution into novel habitats and ways of life such as predatory and aquatic lifestyles. Birds of prey are identified by their powerful beaks and claws. In the case of the claw, studies indicate that the morphology of the claw of birds of prey (also referred to as a talon) differs from non-raptorial birds and between different Orders of birds of prey (Csermely and Rossi, 2006; Fowler et al., 2009) (however see Birn-Jeffery et al., 2012). In addition to being expressed in the claw (Whitbread et al., 1991), claw β-keratins are also expressed in the beak of the chicken (Greenwold and Sawyer, 2010; Wu et al., 2004). In this study we found that the birds of prey have a significantly higher proportion of claw β-keratins than non-raptorial birds, which may

indicate that they have played an important role in the evolution of these unique epidermal appendages.

The feathers of aquatic birds have been shown to have a higher hydrophobicity than the feathers of terrestrial birds (Rijke, 1970). This may be important for thermal regulation especially for birds in adverse climates, such as penguins. Our analysis of β-keratin copy number variation among birds has shown that the proportion of keratinocyte β-keratins is higher and the proportion of feather β-keratins is lower for aquatic birds compared to terrestrial birds. Also, we found that at least 98 of the 133 chicken β-keratin genes are transcribed during the formation of feathers in the chicken. While feather β-keratins are annotated based upon the tissue in which their amino acid sequence was first determined (O'Donnell and Inglis, 1974), it has been shown that there are actually multiple β-keratin gene variants (subfamilies) expressed during embryonic development of feathers (Greenwold and Sawyer, 2013; Presland et al., 1989a and 1989b). Collectively, this indicates that the proportion of gene variants is important as birds have adapted to their lifestyles (aquatic, terrestrial, predatory) and that their feather, claw and beak structure may have been modified by the dynamic expansion and contraction of specific β-keratin gene variants.

CONCLUSION

The number of α-keratin genes is reduced in the avian lineage, and while still important for feather development, for example during rachis morphogenesis (Ng et al., 2012), their low abundance in the barbs and barbules of feathers demonstrates that they have a reduced role in establishing the composition of mature feathers. On the other hand, the β-keratin multigene family has undergone dramatic expansions in the bird and turtle

lineages resulting in novel epidermal appendages (Greenwold and Sawyer, 2010; Greenwold and Sawyer, 2013; Li et al., 2013). Members of all β-keratin subfamilies are expressed during the development of scutate scales and feathers with feather β-keratins becoming specialized in their expression profiles in the diverse assortment of feathers found in present day birds (Kowata et al., 2014; Ng et al., 2014) (Figure 2.4). The early evolution of β-keratins in the archosaurian lineage is marked by lineage specific expansions, but differences in the proportion of claw, feather, and keratinocyte β-keratin genes in modern birds may be attributed to their ecological niche (Figure 2.5). Our overall findings suggest that the number of β-keratins and the relative proportion of β-keratins in each subfamily influenced the composition of avian skin appendages and therefore their structural properties. Clearly, the evolution of feathers in the lineage leading to modern birds has been shaped by the dynamic evolution of α- and β-keratins.

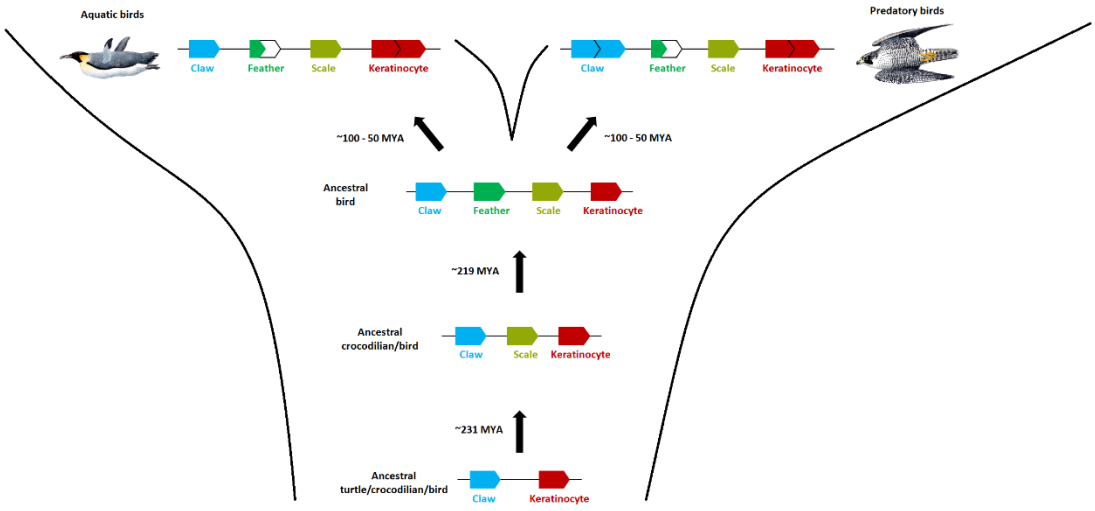


Figure 2.5. Dynamic evolution of β-keratins in the archosaur lineage. This figure illustrates a proposed scenario of β-keratin evolution on the ancestral locus in archosaurs. The bottom row is the proposed locus of β-keratins in the ancestor of turtles, crocodilians and birds. The second row from the bottom indicates that the scale β-keratins have

emerged since the divergence of turtles from crocodilians and birds. The origin of the feather β-keratins occurs after the divergence of crocodilians and birds. The order of the β-keratins subfamilies is based on our genomic data from the 48 birds and green sea turtle. The top row illustrates the dynamic changes of the proportions of β-keratin subfamilies in modern birds with aquatic and predatory lifestyles. Both aquatic and predatory birds have a larger proportion of keratinocyte β-keratins and smaller proportion of feather β-keratins in their genomes. Additionally, predatory birds have a larger proportion of claw β-keratins. The divergence times are in millions of years ago (MYA). The divergence estimates of the turtle – crocodilian/bird split and the crocodilian – bird split are from Shedlock and Edwards (2009). The divergence time estimates for birds is from Jarvis et al. (2014) and is the range starting with the divergence of the Palaeognathae and Neognathae (~100 MYA) and the subsequent divergences of most ordinal groups by ~50 MYA. Jon Fjeldså produced the images of the birds (emperor penguin on the left and peregrine falcon on the right).

METHODS

Genome searches

The genome build information and statistics for the birds used in this study are detailed in Zhang et al. (2014). Additional file 1is reprinted with permission from Zhang et al. (2014) and lists the English names for all species used in this study, while scientific Latin species names are listed in Additional file 2.

Alpha keratin sequences for the chicken, green anole and human were downloaded via NCBI and their accession numbers and names are listed in Additional file 6. The copy number and position of these sequences coincide with the results reported in Vandebergh and Bossuyt (2012). Genome searches were conducted using standalone BLAT (v35) fast sequence search command line tool (Kent, 2002) and GeneWise (v2.2.0) was used as a homology based predictor of gene structure (Birney et al., 2004) with chicken and green anole α-keratin sequences downloaded via NCBI (Additional file 6). Type I and II α-keratin searches were conducted separately. BLAT hits with a Match score, the number of matches minus the mismatches, of ≥ 250 were used to select the best hits for all α-keratin genes. Selected regions were extracted and

gene model structures were predicted using GeneWise (v2.2.0). Bird genomes were searched using chicken α-keratins and the alligator and green sea turtle genomes were searched using both the chicken and green anole α-keratins.

Beta keratin searches of the bird genomes were performed using 18 peer reviewed avian β-keratins sequences from the Swiss-prot database (Additional file 6). These sequences were used to perform a TBLASTN (v2.2.19) search and GeneWise (v2.2.0) was used as a homology based predictor of gene structure (Birney et al., 2004). Avian β-keratins were aligned to the bird genomes using TBLASTN. Aligned regions were selected using the most similar query β-keratin with a length not less than 50% of the protein query. Selected regions were extracted and gene model structures were predicted using GeneWise (v2.2.0). For the green sea turtle genome, we performed a TBLASTN (v2.2.19) search using the Greenwold and Sawyer (2010) data set, which included the crocodilian β-keratins. Following gene selection, we refined and filtered the genes to exclude sequences that did not have an open reading frame (ORF), contained unknown sequence (-NNN-) within the ORF, were less than 80 amino acids, or contained frameshift mutations. The American alligator and green anole β-keratin genes were obtained from Greenwold and Sawyer (2013) and Dalla Valle et al. (2010), respectively.

Gene annotation

Type I and II α-keratin gene annotations were performed using the standalone BLAST+ program and the non-redundant GenBank CDS translations+PDB+SwissProt+PIR+PRF database downloaded February 2013. The β-keratin annotations were performed using the standalone BLAST+ program and the Greenwold and Sawyer (2013) dataset as a database.

Annotation of the avian α-keratins from the 48 bird genomes indicated that several Type I and II α-keratins are not found in birds. In order to verify these findings we downloaded all available EST libraries for the birds used in this study and used the standalone BLAST+ program and tblastn to search the expressed sequence tag (EST) libraries of the chicken (603,076 ESTs), turkey (17,435 ESTs), duck (5,249 ESTs), pigeon (4,931 ESTs) and zebra finch (92,142 ESTs). We identified 484 chicken, 48 turkey, 2 duck, 6 pigeon and 82 zebra finch ESTs as potential α-keratins and annotated them using the above method. Of the potential α-keratin ESTs, 219 chicken, 11 turkey and 6 zebra finch ESTs were annotated as α-keratins while none of the duck or pigeon ESTs were annotated as α-keratins.

Copy number assessment

Species have a shared history that can be represented by a phylogeny and as such, species cannot be considered independent data points in statistical analyses, we therefore employed computer simulations to produce phylogenetic ANOVA p-values in our copy number assessments. As we do not have a species tree that includes the birds, mammals and reptiles used in this study, the statistical analyses of copy number differences for α-keratins were performed by calculating the mean for each group and fold change between the groups.

For the β-keratins, we tested the null hypothesis that the variance of each β-keratin subfamily (claw, feather, scale and keratinocyte) is equal to the variance of the total number of β-keratins using Levene's test and the SPSS software 21 (IBM Corporation, Armonk, NY).

In order to test the null hypothesis that the proportions of β-keratin subfamilies (claw, feather, scale and keratinocyte β-keratins) are the same for all birds, we utilized standard and phylogenetic ANOVA implemented in the R package GEIGER v 2.0.1 (Garland et al., 1993; Harmon et al., 2008) and the species tree from Jarvis et al. (2014) with 10,000 computer simulations. The proportional data used in these analyses were transformed in SPSS using twice the angle (measured in radians) whose trigonometric sine (Arc sine) equals the square root of the proportion being transformed. Furthermore, the transformed data was tested for normality using the lilliefors (Kolmogorov-Smirnov) test in SPSS and found to be normally distributed. Following phylogenetic ANOVA analyses, the residuals were analyzed and found to be normally distributed with the lilliefors (Kolmogorov-Smirnov) test implemented using the R package nortest.

For the phylogenetic ANOVA analyses, we grouped birds according to their phylogenetic relatedness or lifestyle (aquatic or predatory). For species phylogenetic relatedness, we performed two analyses. The first grouped Paleognathae (ostrich and tinamou) compared to all other birds (Neognathae), the second grouped Paleognathae and Galloanserae (chicken, turkey and duck) compared to all other birds (see Additional file 3; [Jarvis et al., 2014]). For the first lifestyle comparison, we grouped the aquatic and semi-aquatic birds (See Additional file 1; [Zhang et al., 2014], Additional file 3; [Jarvis et al., 2014]) and compared them to the "land" birds (the terrestrial birds). As the eagles do not cluster with other aquatic birds (Additional file 3; [Jarvis et al., 2014]), we also performed the aquatic lifestyle comparison with the eagles grouped with land birds. The final lifestyle comparison analyzed if birds with a predatory lifestyle have any significant

differences when compared to the remaining birds (Additional file 1; [Zhang et al., 2014], Additional file 3; [Jarvis et al., 2014]).

Phylogenetic analyses

All phylogenetic analyses were performed using RAxML version 7.4.4 with 1000 bootstrap replicates and the rapid bootstrap analysis (Stamatakis, 2006). Amino acid substitution models were selected using Prottest version 3.2 (Abascal et al., 2005) and amino acid alignments were performed using ClustalW 2.0.10 (Larkin et al., 2007).

The α-keratin Type I dataset consisted of a total of 580 sequences and the Type II dataset consisted of 625 sequences obtained from the genomes of 48 birds, the American alligator and green sea turtle (St John et al., 2012; Wang et al., 2013), while the sequences used for the green anole and human were downloaded via NCBI (Additional file 6). The β-keratin phylogeny (Additional file 4; [Zhang et al., 2014]) was performed using all 1,623 complete bird genes and the β-keratins from three non-avian reptiles; the green anole lizard (Dalla Valle et al., 2010), green sea turtle (Li et al., 2013) and American alligator (Greenwold and Sawyer, 2013) for a total of 1,698 β-keratins.

Previous molecular analyses of α-keratins indicated that Type I and II sequences should not be analyzed together because the high degree of sequence divergence between the two types result in non-informative phylogenies (Vandebergh and Bossuyt, 2012). Only amino acid sequences containing the entire central helical domain (rod domain) were used in phylogenetic analyses. For the Type II α-keratins this eliminated the 46 KRT6A avian genes.

The Prottest program found that the JTT model with a gamma distribution as the best fit model under the Akaike Information Criterion and Bayesian Information

Criterion for both Type I and II α-keratins. Model selection for the β-keratins found that the WAG amino acid model with a gamma distribution as the best fit model under the Akaike Information Criterion and Bayesian Information Criterion for β-keratins.

Genomic orientation figures

For Type I and II α-keratins, we constructed a proposed genomic orientation based on all of the 48 birds, of which each had at least one genomic locus with at least 2 or more gene variants for both Type I and II α-keratins. Although the results presented in Figure 2.1B and 2.2B is representative of most birds, some bird species have additional gene duplications not displayed in the figures. For Type I α-keratins, the red-legged seriema has an additional KRT12 next to the KRT12 on Figure 2.1B. Type II α-keratins showed more variation. Five birds (MacQueen's bustard, cuckoo-roller, hoatzin, red-crested turaco and barn owl) have an additional KRT75 between clade C1 and C2 in Figure 2.2B. Also, four birds (golden-collared manakin, white-tailed tropicbird, downy woodpecker and great-crested grebe) have an additional KRT5 between clade C1 and C2. Two species (Macqueen's bustard and cuckoo-roller) have an additional cochleal gene in clade C2. Finally, the white-tailed eagle has an additional KRT84 between clade C1 and C2 and the white-tailed tropicbird has an additional KRT73 next to the other KRT73 in clade B1. The annotation of the gene variants is based upon the molecular phylogenies in Figure 2.1A and 2.2A.

For the β-keratins, we constructed a genomic alignment of the β-keratin locus containing all of the β-keratin subfamilies (microchromosome 25 of the chicken and zebra finch) for birds with at least one genomic locus with at least 2 β-keratin subfamilies so we could ascertain the orientation of each genomic locus relative to the other birds. A

total of 20 phylogenetically diverse species are included in Figure 2.3, which comprise closely related species pairs (penguins and finches) that illustrate genomic conservation at the fine and broad taxonomic scale across birds. The line breaks for each species is indicative of different genomic scaffolds or contigs. Even though species may lack a region or genes of a β-keratin subfamily (such as the duck and clade 2 claw β-keratins) does not mean they do not have clade 2 claw β-keratins, but that the species does not have a genomic locus with at least 2 subfamilies covering that area. Furthermore, the genes were aligned based on gene orientation and to the species with the highest number of genes similar to DNA sequence alignments with indels. Therefore, species such as the kea and pigeon appear to have a large gap for the scale β-keratins when in fact the genomic region was extended for the chicken which has the highest number of scale β-keratins. Arrows indicate the transcriptional direction. Incomplete β-keratin genes are indicated by white filled arrows, while the solid colored arrows indicate complete genes. The annotation of the β-keratin genes is based upon our annotation (above the figure) and phylogenetic analyses (below the figure).

Differential α- and β-keratin expression in chicken epidermal tissues during embryogenesis

We investigated the expression profiles of α- and β-keratins during feather and scale development in the chicken using a customized version of the chicken 44K Agilent microarray (Li et al., 2008). This 60-mer oligonucleotide microarray was validated and tested on a diverse array of tissues and was found to have high sensitivity and specificity (Li et al., 2008). In order to include α-keratins and β-keratins, we removed genes that we deemed not important in this study, such as sensory receptor genes, avian viruses,

immune genes and genes related to vision. We were able to create unique 60-mer oligonucleotide sequences for all 27 α-keratins and 102 β-keratins using the online tool, Agilent's eArray application (Agilent Technologies; Palo Alto, CA).

Macro tissue dissections from the chicken dorsal feather and scutate scale were taken at day 8, 17 and 19 of embryonic development, while the wing feather dissections were taken at day 17 and 19. All tissues, immediately following dissection, were fixed in RNAlater (Qiagen; Germantown, MD). Chicken tissue samples were kindly provided by Richard Goodwin at the University Of South Carolina School Of Medicine in Columbia, SC. RNA extractions were performed using the Qiagen miRNeasy Mini Kit (Qiagen; Germantown, MD). RNA quality and quantity were checked by both the Agilent Technologies 2100 Bioanalyzer (Agilent Technologies; Palo Alto, CA) and Thermo Scientific NanoDrop 2000c spectrometer (Thermo Scientific; Waltham, MA).

RNA was converted to cDNA, labeled and hybridized to the aforementioned single color custom Agilent microarray gene chip at the South Carolina College of Pharmacy Microarray Core Facility (University of South Carolina, Columbia, S.C.). All samples were analyzed in replicates of four, except for day 8 scutate scale which was analyzed in triplicate. The 8 samples were used to perform 14 sample comparisons (See Table 2.2 and Additional file 5). Both clustering and sample comparisons were performed using Agilent GeneSpring software12.5 (Agilent Technologies; Palo Alto, CA). Sample comparisons were performed using the Mann-Whitney U-test with a p-value cut-off of 0.05 and fold change cut-off of 2.0.

Availability of supporting data

The data sets supporting the results of this study can be found at Beijing Genomics Institute (http://phybirds.genomics.org.cn) and the GigaScience Genome Downloads (http://gigadb.org/dataset/101000) [Zhang et al., 2014; Jarvis et al., 2014].

CHAPTER 3

EXPRESSED miRNAs TARGET FEATHER RELATED mRNAs INVOLVED IN CELL SIGNALING, CELL ADHESION AND STRUCTURE DURING CHICKEN EPIDERMAL DEVELOPMENT

Bao, W., Greenwold, M. J., Sawyer, R. H. (2016). *Gene*, 591 (2), 393-402

ABSTRACT

MicroRNAs (miRNAs) are small non-coding RNAs that regulate gene expression at the post-transcriptional level. Previous studies have shown that miRNA regulation contributes to a diverse set of processes including cellular differentiation and morphogenesis which leads to the creation of different cell types in multicellular organisms and is thus key to animal development. Feathers are one of the most distinctive features of extant birds and are important for multiple functions including flight, thermal regulation, and sexual selection. However, the role of miRNAs in feather development has been woefully understudied despite the identification of cell signaling pathways, cell adhesion molecules and structural genes involved in feather development. In this study, we performed a microarray experiment comparing the expression of miRNAs and mRNAs among three embryonic stages of development and two tissues (scutate scale and feather) of the chicken. We combined this expression data with miRNA target prediction tools and a curated list of feather related genes to produce a set of 19 miRNA-mRNA duplexes. These targeted mRNAs have been previously identified as important cell signaling and cell adhesion genes as well as structural genes involved in feather and scale morphogenesis. Interestingly, the miRNA target site of the cell signaling pathway gene, Aldehyde Dehydrogenase 1 Family, Member A3 (ALDH1A3), is unique to birds indicating a novel role in Aves. The identified miRNA target site of the cell adhesion gene, Tenascin C (TNC), is only found in specific chicken TNC splice variants that are differentially expressed in developing scutate scale and feather tissue indicating an important role of miRNA regulation in epidermal differentiation. Additionally, we found that β-keratins, a major structural component of avian and reptilian epidermal

appendages, are targeted by multiple miRNA genes. In conclusion, our work provides quantitative expression data on miRNAs and mRNAs during feather and scale development and has produced a highly diverse, but manageable list of miRNA-mRNA duplexes for future validation experiments.

INTRODUCTION

Regulation of gene expression contributes to phenotypic differences between different structures (Lockhart and Winzeler, 2000). Although many other mechanisms such as alternative splicing and polyadenylation are involved in modulating the expression of genes on the mRNA level (Chen and Rajewsky, 2007), microRNA (miRNA) regulation has been recognized as one of the key mechanisms in the mRNA transcript regulatory network in eukaryotes (He and Hannon, 2004). MicroRNAs are small endogenous noncoding RNAs between 19-24 nucleotides (nt) in size (Bartel, 2004; Carrington and Ambros, 2003). A mature miRNA is generated from a pre-miRNA, also referred to as a miRNA gene (Kamanu et al., 2013), which is a ~70 nt hairpin precursor. Mature miRNAs can complementarily bind target mRNA sequences to inhibit translational initiation or lead to mRNA degradation. In animals, it is reported that miRNAs can bind with the coding sequence region (CDS), 3' or 5' untranslated regions (UTR) of target mRNAs (Duursma et al., 2008; Lytle et al., 2007; Wu et al., 2006).

MicroRNAs have been found to widely regulate many biological processes such as development, metabolism, organogenesis, differentiation and proliferation in animals (Betel et al., 2008) including the spatio-temporal patterns important for development (Miska et al., 2004). The chicken (Gallus gallus) embryo develops and hatches in just ~20-21 days (Hamburger and Hamilton, 1951) and in this short time span, many unique

and fascinating structures such as scales and feathers are formed during this complex and amazing transformation from an embryo into chick. Only 23 miRNA target genes have been validated through experimentation in the chicken (Hsu et al., 2014) and none of these are related to chicken embryonic development. However, Zhang et al. (2013) conducted a miRNA profiling analysis of the duck feather follicle and skin using high-throughput sequencing data. Using only miRNA differential expression data and only genes involved in a few signaling pathways (Wnt/β-catenin, SHH-BMP2 and Notch signal pathways) for miRNA target prediction, they found that seven miRNAs are involved in feather morphogenesis. However, there are more than 700 predicted miRNAs in the chicken and each miRNA may target multiple mRNAs at multiple sites (Lewis et al., 2005) thereby increasing the complexity of the miRNA regulation network. Therefore, identification of miRNA targets coupled with mRNA and miRNA expression data is critical to advancing our understanding of the function of miRNAs during chicken epidermal appendage development.

Recently, it has been found that a high number of β-keratins are expressed during the development of scutate scales and feathers in the chicken using high-throughput microarray technology (Greenwold et al., 2014; Ng et al., 2014). However, at the protein level, only a few β-keratin species can be detected from scale and feather tissues on 2-dimensional gels at the same developmental stages (Knapp et al., 1991; Knapp et al., 1993; Rice et al., 2013; Shames and Sawyer, 1986; Shames et al., 1988), suggesting that β-keratin mRNAs are being regulated possibly by miRNAs.

This study profiled differentially expressed chicken mRNA and miRNA genes among 14 comparison groups from three important embryonic stages of development

(day 8, 17 and 19) and between two tissue types (scutate scale and feather). Although feather development begins as early as day 6, the formation of barbs and barbules does not begin until ~day 12 of embryogenesis (Lucas and Stettenheim, 1972; Sengel, 1976) and in the case of scutate scales, epidermal placodes do not begin to develop until day 9 of embryogenesis making day 8 of embryogenesis an ideal control for feather and scale tissue (Shames and Sawyer, 1986). The scutate scale of the chicken abundantly expresses β-keratin at embryonic day 17 (Shames and Sawyer, 1986). Therefore, day 8 is the stage before epidermal tissue development, day 17 is a peak stage of epidermal development and day 19 is a late development stage close to hatching (~20-21 days).

In order to study the role of miRNAs in the development of scutate scales and feathers, miRNA target prediction tools were customized to produce a list of top ranked differentially expressed target mRNA genes from the differentially expressed miRNA genes of our microarray experiment. These identified target genes are known as important cell signaling (Fibroblast Growth Factor 20 [FGF20] and Aldehyde Dehydrogenase 1 Family, Member A3 [ALDH1A3]), cell adhesion (Cadherin 1, type 1 [CDH1] and Tenascin C [TNC]) and structural genes (α- and β-keratins) involved in avian epidermal appendage morphogenesis. Cell signaling molecules are important for gross tissue patterning, cell adhesion molecules are important in the intercellular interactions that result in the complex pattering of the feather rachis, barbs and barbule components while the structural proteins lend unique biomechanical properties to scales and feathers (Fraser and Parry, 1996; Lin et al., 2006; Maderson et al., 2009; Pabisch et al., 2010; Weiss and Kirchner, 2011). Therefore, our findings provide a mechanism of miRNA regulation during embryonic scutate scale and feather differentiation.

MATERIALS AND METHODS

RNA isolation and purification

Dorsal feather (DF) and scutate scale (SC) tissues were taken at day 8, 17 and 19 of chicken embryonic development, while the wing feather (WF) tissue was taken at day 17 and 19. The tissue samples were of the whole skin which includes the epidermis and dermis. All tissues after the dissection were fixed immediately in RNAlater (Qiagen; Germantown, MD). Chicken tissue samples were generously provided by Dr. Richard Goodwin at the School of Medicine of the University of South Carolina. Total RNA was isolated by using Trizol (Life Technologies; Carlsbad, CA) and Qiagen miRNeasy Mini Kit (Qiagen; Germantown, MD) with the manufacturer's instructions. The quantity and quality of extracted RNA were evaluated by both the Agilent Technologies 2100 Bioanalyzer (Agilent Technologies; Palo Alto, CA) and Thermo Scientific NanoDrop 2000c spectrometer (Thermo Scientific; Waltham, MA).

Microarray Experiment

A customized version of the chicken 44K Agilent microarray was used in this study. This 60-mer oligonucleotide microarray has been validated and tested in different tissues and demonstrated to have high sensitivity and specificity (Li et al., 2008). We removed non-relevant genes such as avian viruses, immune genes, and sensory receptor genes. Using Agilent's (Agilent Technologies; Palo Alto, CA) eArray web based tool (http://www.genomics.agilent.com/en/Custom-Design-Tools/eArray), we created unique 60-mer oligonucleotide sequences for 639 chicken specific stem loop pre-miRNAs (miRNA genes) on July 10th 2012, as well as 27 α-keratins and 102 β-keratins for the microarray.

Total RNA was reversed transcribed into cDNA and then the cDNA was fragmented, labelled and hybridized to the single color custom Agilent microarray gene chip at the South Carolina College of Pharmacy Microarray Core Facility (University of South Carolina, Columbia, SC). After hybridization and washing, the processed slides were scanned with the Agilent High-Resolution Microarray Scanner (Agilent Technologies; Palo Alto, CA). All 8 samples were analyzed with 4 biological replicates, except for the scutate scale day 8 sample which was in triplicate thereby resulting in a total of 31 individual microarray expression samples.

Raw gene expression level files were imported into Agilent GeneSpring software 12.5 (Agilent Technologies; Palo Alto, CA) and normalized by the quantile method, enabling 14 sample comparisons with the 8 samples. Hierarchal clustering with Euclidean distance and normalized intensity values of mRNA genes was also conducted by GeneSpring to analyze overall sample relationships (Figure 3.1b). Differentially expressed genes were determined using the Mann-Whitney U-test with p-values ≤ 0.05 and the fold change expression ≥ 2. We also performed a hierarchical cluster analysis for the 31 individual microarray expression samples using the R package pvclust (Suzuki and Shimodaira, 2006) and the Approximately Unbiased (AU) p-value and Bootstrap Probability (Additional File 3.1).

Kane et al. (2000) found that 50 mer microarray oligonucleotide probes will cross hybridize with non-specific targets with an equal to or greater similarity of 80%. Therefore, we performed a BLASTN search (Basic Local Alignment Search Tool nucleotide-nucleotide v. 2.2.23+; Altschul et al., 1990) of the chicken genome (v. 4.0) with the 226 differentially expressed 60 mer oligonucleotide miRNA probes. We found

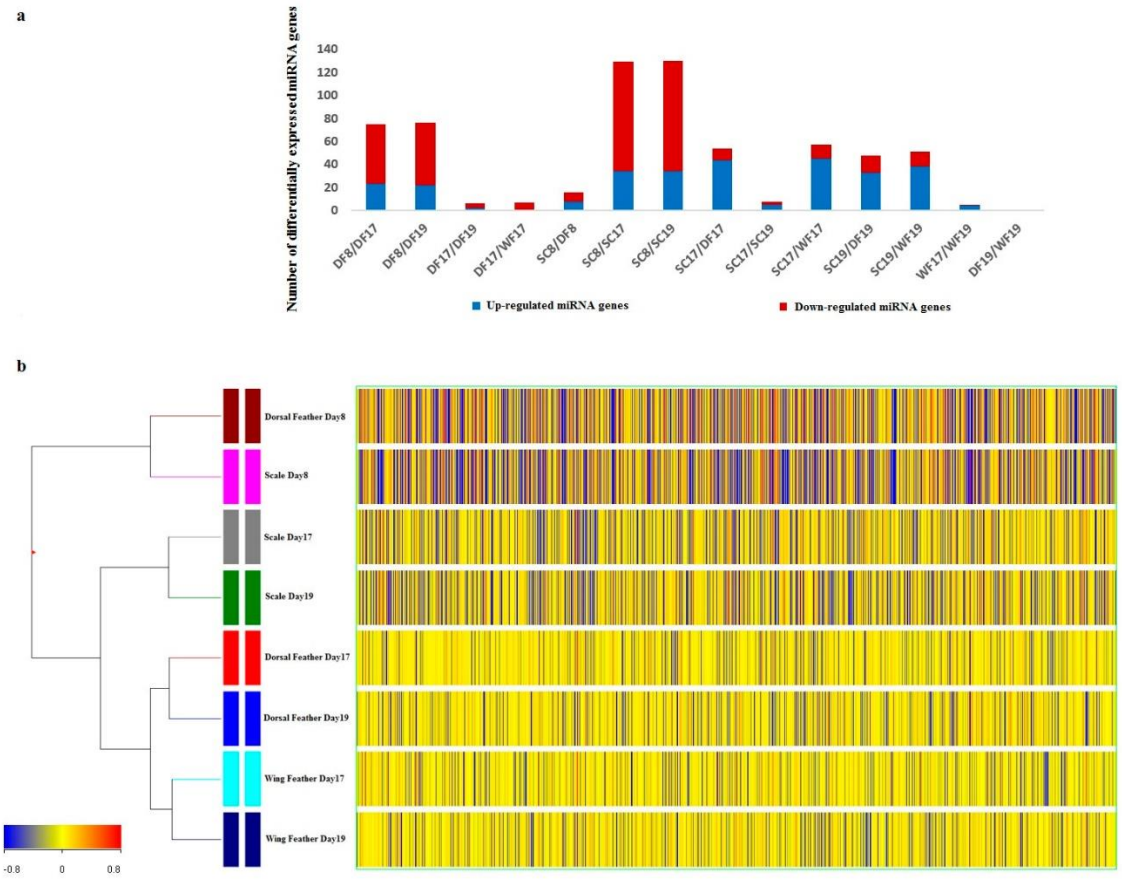


Figure 3.1. (a) Graphically displays the number of differentially expressed miRNA genes for all 14 comparison groups. (b) Illustrates a hierarchical clustering dendrogram of the biological replicates grouped by their respective microarray sample and based on the normalized intensity values of mRNA genes using the Euclidean similarity measure.

that ~93% (209 of 226) of the probes had only one unique hit against the chicken genome indicating that the majority of our differentially expressed miRNAs expression values were specific to the intended miRNA gene.

Genome searches and sequence alignments

737 total chicken miRNA gene sequences (miRNA primary transcript) were retrieved from miRBase database (Griffiths-Jones et al., 2008). A BLASTN (Basic Local Alignment Search Tool nucleotide-nucleotide v. 2.2.23+; Altschul et al., 1990) search

with default settings was conducted using these 737 chicken miRNA genes as query for the genomes of an amphibian (Xenopus Silurana tropicalis) (Hellsten et al., 2010), three mammals (Ornithorhynchus anatinus, Mus musculus and Homo sapiens) (1000 Genomes Project Consortium, 2012; Warren et al., 2008; Takada et al., 2013), one lizard (Anolis carolinesis) (Alföldi et al., 2011), three turtles (Chelonia mydas, Chrysemys picta belli and Pelodiscus sinensis) (Shaffer et al., 2013; Wang et al., 2013), three crocodilians (Alligator mississippiensis, Gavialis gangericus and Crocodylus porosus) (St John et al., 2012), one snake (Burmese python) (Castoe et al., 2013) and 48 birds (Zhang et al., 2014). The sequencing depth of the 45 bird genomes are grouped into high (>50X) and low coverage (~30X) assemblies while the chicken, zebra finch and turkey genomes are robust builds with the majority of the contigs/scaffolds anchored to chromosomes. While this variation may imply significant differences in genome coverage, genome sizes of all 45 birds are nearly complete based on cytology based genome sizes (Zhang et al., 2014).

The BLAST results were manually inspected to remove poor hits. BLAST hits are either a full length hit with a BLAST Score ≥ 51 bits, e-value ≤ 4e-0.005 and Identity ≥ 77% or resulted in no hits. These search methods were further validated using a reciprocal BLAST search of the chicken miRNA genes against the chicken genome. We found that 4 miRNA genes (gga-mir-466, gga-mir-1552, gga-mir-1763 and gga-mir-193a) in miRBase lack BLAST hits in the chicken genome (assembly Gallus_gallus-4.0). This is due to these 4 genes being consigned to the obsolete NCBI Trace Archive. However, we still used these genes as queries to search the 59 non-chicken genomes. We also observed that some chicken miRNA genes have multiple hits in the chicken genome. For example, gga-mir-1594 has 61 hits on 33 scaffolds of chicken genome.

Full length nucleotide sequence alignments were performed with default settings using ClustalW (Larkin et al., 2007). The alignment viewer, AliView (Larsson, 2014), was applied to illustrate the conservation of target site sequences.

miRNA target prediction

MicroRNA genes are the precursors of mature miRNAs and each miRNA gene can form one or two mature miRNAs (Bartel, 2004). From the 226 differentially expressed miRNA genes from our microarray experiment, we retrieved 310 mature chicken miRNAs from the miRBase. These mature miRNA sequences were used as queries for computational miRNA target prediction.

A chicken messenger RNA (mRNA) database was created by downloading all of the chicken (Gallus gallus) mRNAs available from NCBI on May 28th, 2014 and adding 34 α-keratins and 133 β-keratins nucleotide sequences from Greenwold et al. (2014) due to the lack of α- and β-keratin mRNAs. This resulted in a database consisting of 61,408 entries.

RNAhybrid is a miRNA target prediction tool which finds energetically favorable hybridization sites of a miRNA in a large data set of target mRNA candidates and reports the minimum free energy (mfe) value to yield a secondary RNA duplex pairing model (Rehmsmeier et al., 2004). Furthermore RNAhybrid, using Poisson statistics, estimates a p-value for each miRNA-mRNA binding site by modeling multiple binding sites for each mRNA target (Rehmsmeier, 2006). Scale parameters of the extreme value distribution were calibrated using RNAcalibrate for each miRNA prior to running RNAhybrid. RNAhybrid was subsequently applied to model multiple binding sites for each differentially expressed miRNA against our mRNA database (see above). Predicted

binding sites can be either in the 3' and 5' region (UTR) or coding (CD) region (Duursma et al., 2008; Lytle et al., 2007; Wu et al., 2006), therefore we did not trim the mRNA sequences in our database. This resulted in each of the 310 mature miRNAs of the 226 differentially expressed miRNA genes in each comparison group having multiple target mRNAs with mfe values and p-values. We ranked the 622 resulting mRNA targets with a p-value threshold of 0.01 based on the highest mfe value. MicroRNA-mRNA duplexes with lower mfe values are more stable, which means they have a higher chance of binding that specific site of a specific mRNA. We retrieved the binding sites of the mRNA targets from the RNAhybrid output file to study the conservation of the target sites among amniotes.

A recent study by Lowe et al. (2015) conducted a literature search for genes related to feather development. Lowe et al. (2015) found 193 keratin and non-keratin feather-related genes (see Supplemental Table 1 in Lowe et al., 2015). We removed the keratin feather-related genes and replaced them with the α- and β-keratins annotated and localized in the chicken genome from Greenwold et al. (2014) resulting in a final list of 302 feather-related genes. While the Lowe et al. (2015) study focused on feather-related genes, it is likely that many of these genes are also utilized in avian scutate scale development (Prin and Dhouailly, 2004). Using the 302 feather-related genes (see Supplemental Table 1 in Lowe et al., 2015), we sought to minimize the 622 possible miRNA target genes (see above) to a more manageable list. Therefore, we cross compared the 622 target genes with the list of the 302 curated keratin and non-keratin feather-related genes. This resulted in 21 target genes distributed among 8 of the microarray comparison groups (see Table 3.1). All 60 mer microarray probes for the 14

miRNAs listed in Table 3.1 had one unique hit to the chicken genome (see Material and Methods above) suggesting that the expression of these miRNAs are not associated with non-specific binding.

Table 3.1. Differentially expressed miRNAs and their differentially expressed target mRNA genes with the top minimum free energy value which are also classified as feather-related genes by Lowe et al. (2015).

Comparison Group	miRNA	Origin of miRNA[a]	Target gene	Target binding
SC8/DF8	gga-miR-10b-3p	Amniota	Aldh1a3	Coding region
SC8/DF8	gga-miR-205a	Tetrapoda	GALGA_5AKI_KRT19 [c]	Coding region
SC8/SC17	gga-miR-6606-5p	Aves	Epha4	5' UTR
SC8/SC17	gga-miR-6651-5p	Galliformes	Fgf20	Coding region
SC8/SC19	gga-miR-3538	Aves	Cdh1	Coding region
SC8/SC19	gga-miR-6651-5p	Aves	Fgf20	Coding region
SC17/DF17	gga-miR-6679-5p	chicken	Edar	Coding region
SC17/DF17	gga-miR-200b-5p	Tetrapoda	Slc45a2	Coding region
SC17/DF17	gga-miR-146a-5p	Amniota	TNC	Coding region
SC19/DF19	gga-miR-1759-3p	Aves	Gbx2	Coding region
SC19/DF19	gga-miR-1463	Galliformes	GALGA25_Ktn11 [b]	Coding region
SC19/DF19	gga-miR-6660-3p	chicken	GALGA_5AKII_KRT5 [c]	Coding region
SC19/WF19	gga-miR-1759-3p	Aves	Gbx2	Coding region
WF17/WF19	gga-miR-1806	Galliformes	GALGA_UnR_FK10 [b]	Coding region
DF17/DF19	gga-miR-1556	chicken	GALGA25_CL1 [b]	Coding region
DF17/DF19	gga-miR-1556	chicken	GALGA25_CL2 [b]	Coding region

DF17/DF19	gga-miR-1556	chicken	GALGA25_CL6 [b]	Coding region
DF17/DF19	gga-miR-1556	chicken	GALGA25_CL7 [b]	Coding region
DF17/DF19	gga-miR-1556	chicken	GALGA25_CL8 [b]	Coding region
DF17/DF19	gga-miR-1556	chicken	GALGA_UnR_CL1 [b]	Coding region
DF17/DF19	gga-miR-204	Amniota	GALGA25_Ktn8 [b]	Coding region

a: Origin of miRNAs is in reference to Figure 3.2;
b: β-keratin genes and the number after GALGA represent the genomic chromosome loci of the gene and UnR represents unknown location; CL and Ktn represent claw β-keratin and keratinocyte β-keratin respectively.
c: α-keratin genes. AKI and AKII represent Type I (acidic) and Type II (basic) α-keratin respectively.

The differentially expressed miRNAs listed in Table 3.1 were validated using RNA22 (Miranda et al., 2006), which searches the whole target mRNA. RNA22 settings were a seed size of 8, maximum of 2 unpaired bases, minimum number of paired-up bases for heteroduplexes is 10, maximum folding energy for heteroduplexes is -5 Kcal/mol and a threshold for searching sensitivity/specificity of 500,000.

We investigated the conservation of the miRNA predicted target site of the target genes (Table 3.1) among representative species of tetrapod groups in Figure 3.2. Each non-keratin target gene listed in Table 3.1 was downloaded from the NCBI mRNA database and all incomplete genes were manually removed. We then used the chicken's predicted target site from the RNAhybrid results for each miRNA gene as a query for standalone BLASTN searches (Basic Local Alignment Search Tool nucleotide-nucleotide v. 2.2.23+; Altschul et al., 1990) against all available mRNA genes downloaded from NCBI using a BLAST score and e-value cutoff of 30 and 0.005, respectively.

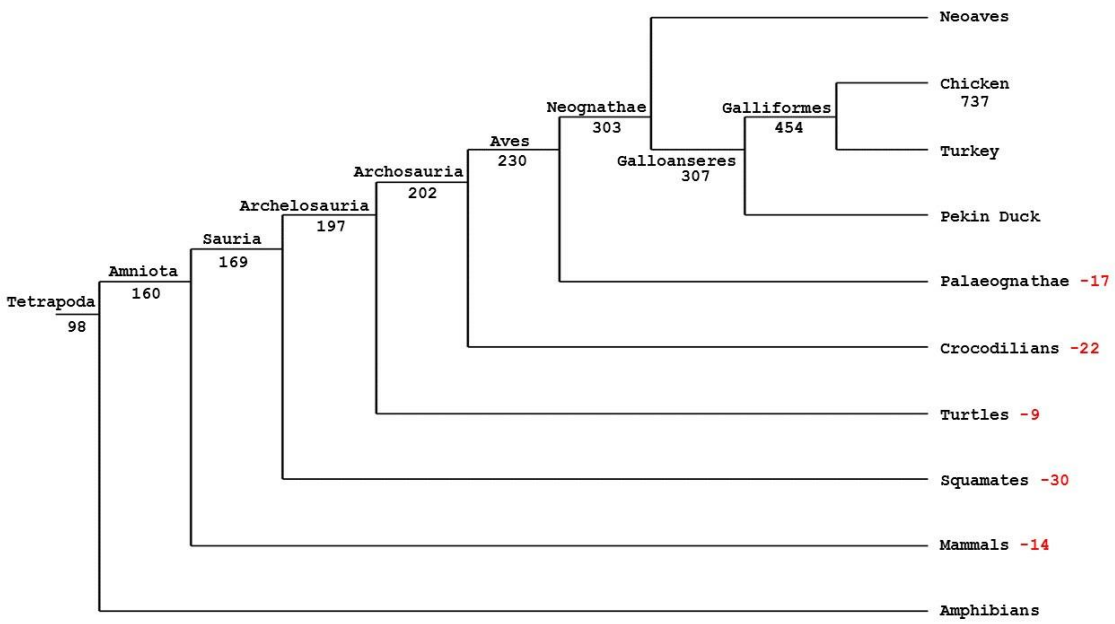


Figure 3.2. Displays the probable birth of chicken miRNA genes across the tetrapod phylogeny (Jarvis et al. 2014; Crawford et al. 2015) based on the presence or absence of the miRNA chicken genes in genome searches (see also Additional File 3.2 for detailed searched results). The numbers for each broad taxonomic classification (Tetrapoda, Amniota, Sauria etc.) are the accumulative number of chicken miRNA genes found in the most recent common ancestor. The red number next to the clade names are the number of miRNA genes that are not found in that clade but are found in the hierarchical most recent common ancestor.

RESULTS

Identification of miRNAs involved in epidermal development

We performed comparative transcriptome analyses on tissue from the chicken scutate scale (SC), dorsal feather (DF) and wing feather (WF) at day 17 and 19 of embryogenesis and the SC and DF tissue at day 8 of embryogenesis using a customized version of the chicken 44K Agilent microarray with 639 miRNA genes and 30900 total genes. A total of 226 miRNA genes were found to be differentially expressed in 13 out of 14 comparison groups (Figure 3.1a; Additional File 3.3). Comparison of DF19 and WF19 did not result in any differentially expressed miRNA genes with comparison group

SC8/SC19 resulting in the largest number (130) of differentially expressed miRNA genes (Figure 3.1a).

SC8/SC17 and SC8/SC19 have 129 and 130 differentially expressed miRNA genes, respectively, while SC17/SC19 only has 8 differentially expressed miRNA genes. Similarly, DF8/DF17 and DF8/DF19 have a large number of differentially expressed miRNA genes while DF17/DF19 and WF17/WF19 comparison groups have a much reduced number of differentially expressed miRNA genes. Therefore, the day 8 to day 17 and 19 comparisons for both the scutate scale and dorsal feather have a larger number of differentially expressed miRNA genes than intra-tissue comparisons between day 17 and 19 indicating that there is a temporal effect on miRNA gene expression.

When comparing scutate scale tissue with feather tissues (inter-tissue) from the late development stages such as SC17/DF17, SC17/WF17, SC19/DF19, SC19/WF19, the number of differentially expressed miRNA genes range from ~48-57 which is larger than the number found in the early versus late development stage comparison group (SC8/DF8; 17 differentially expressed miRNA genes) and in the different feather inter-tissue comparison groups such as DF17/WF17 (7 differentially expressed miRNA genes) and DF19/WF19 (0). These results may indicate that the set of miRNAs used during late scale development (day 17 and 19) are different compared to late feather development, but that the two feather tissues (DF and WF) utilize a similar set of miRNA genes expressed at nearly identical levels. Furthermore, different miRNA genes are differentially expressed temporally during embryonic development and thus may be important for the specific tissue development which is congruent with the number of

differentially expressed miRNA genes of the 14 comparison groups in Figure 3.1a (Additional File 3.4).

Hierarchical clustering of the mRNA genes grouped by microarray tissue sample displayed in Figure 3.1b illustrates that there are three major clades in the hierarchical clustering; an early development (day 8) cluster and two clusters of late development based on tissue (scutate scale and feather) indicating that embryonic epidermal tissue becomes more specialized with maturation. Hierarchical clustering was validated with bootstrap resampling of microarray biological replicates using R pvclust package (Suzuki and Shimodaira, 2006) (Additional File 3.1). Similar to mRNA clustering, miRNA gene expression indicates that the greatest number of differentially expressed RNA (mRNA and miRNA) genes is found between day 8 and day 17/19 intra-tissue comparison groups further supporting the significant impact of embryonic maturation on epidermal gene expression.

Origin of chicken miRNA genes across Tetrapods

Feathers are unique to extant birds and since this study's main focus was to explore the possible involvement of miRNAs in the development of avian epidermal appendages, we investigated whether differentially expressed miRNA genes of our transcriptome experiment may be an innovation unique to birds as are their feathers. There are only 2 avian species in the current miRBase miRNA database (Griffiths-Jones et al., 2008), therefore, we searched 60 Tetrapod genomes including one amphibian, 3 mammals, 2 squamates, 3 turtles, 3 crocodilians and 48 birds (see Material and Methods for a detailed list) using the 737 chicken miRNA genes currently available in miRBase. Figure 3.2 displays the probable birth of chicken miRNA genes across tetrapods (see also Additional

File 3.2 for details). We found that over 13% of the chicken miRNA genes were likely present in the most recent common ancestor (MRCA) of amphibians and amniotes, and that nearly ~22% of the chicken miRNA genes are present in the MRCA of amniotes indicating that either the chicken has retained a large number of amniote miRNA genes vital for organismal function or more likely that a burst in the number of miRNA genes occurred during the evolution of amniotes.

Since the amniote explosion of miRNA genes, a steady accumulation of miRNAs occurred until the diversification of Galliformes from other Galloanseres (Figure 3.2) where the proportion of chicken miRNA genes jumped from ~42% to ~62%. In fact, we found that only 31% of the chicken miRNA genes are conserved across birds. Within birds, the closer related the species are, the larger the number of specific miRNA genes shared because although the turkey and chicken are Galliformes and share ~62% of the chicken miRNA genes, there are still 283 chicken specific miRNA genes which is consistent with previous reports that there are many species specific miRNAs in animals (Gardner et al., 2015; Hoeppner et al., 2012).

Interestingly, we found that 148 chicken miRNA genes are only found in Galliformes (chicken and turkey) and 2 genes are only found in Galloanseres (Additional File 3.2). It is notable that only one chicken miRNA gene (gga-mir-3535) was found in all 48 birds but not any other non-avian species. All of the other avian specific miRNA genes (230) are randomly distributed based on the avian phylogeny (Jarvis et al., 2014) and there is no evidence that the distribution of chicken miRNAs is correlated to phylogenetic relatedness or avian lifestyles (Jarvis et al., 2014; Zhang et al., 2014).

Computational identification of differentially expressed miRNA-mRNA duplexes

Table 3.1 lists the highly probable miRNA-mRNA duplexes involved in epidermal differentiation identified using RNAhybrid algorithm (Rehmsmeier et al., 2004) and our microarray expression data of miRNAs and mRNAs (See Material and Methods and Additional File 3.5 for details). Of the 19 different mRNA target genes listed in Table 3.1, Fibroblast Growth Factor 20 (FGF20) was found to be targeted by the same miRNA in two comparison groups and Gastrulation Brain Homeobox 2 (GBX2) was also found in two different comparison groups. Furthermore, 1 feather β-keratin, 2 α-keratins, 2 keratinocyte β-keratins are targeted by different miRNAs and 6 claw β-keratins are targeted by the same miRNA (Table 3.1). The RNA22 target prediction algorithm (Miranda et al., 2006) also verified the miRNA-mRNA duplexes in Table 3.1. The RNA 22 algorithm predicted the same target site for EPHA4, FGF20, CDH1, EDAR, GBX2 and all Claw β-keratins as RNAhybrid.

We hypothesized that if the miRNA targeting and regulation of genes found in Table 3.1 are associated with the evolution of feathers, then the miRNA gene or mRNA target site may be conserved in birds to the exclusion of other tetrapods. We first determined when these miRNA genes originated during the evolution of tetrapods in reference to Figure 3.2. We found that two miRNA genes are found in all tetrapods, three miRNA genes are found in all amniotes, five miRNA genes originated in birds and nine miRNA genes are only found in Galliformes or the chicken (Table 3.1).

Next, we investigated the conservation of the miRNA target sites in their target genes in combination with the miRNAs' presence (Table 3.1) among tetrapod species (Figure 3.3; also see Material and Methods). Interestingly, 47 of the 186 complete Aldehyde Dehydrogenase 1 Family, Member A3 (ALDH1A3) genes downloaded from

NCBI that had a BLAST hit for the conserved target site belonging to 37 different species of birds (see Additional File 3.6 for details). However, 37 of these 47 ALDH1A3 sequences belonging to 30 birds have the same two nucleotide mismatches in the target site. The conservation of the ALDH1A3 target site in birds compared to other tetrapods is illustrated in Figure 3.3a. Furthermore, we performed a RNAhybrid analysis with gga-miR-10b-3p and the ALDH1A3 genes displayed in Figure 3.3a and found that the target site in non-avian ALDH1A3 genes is different from that of birds and that the minimum free energy (mfe) value for non-avian species ranged from -19.3 to -23.6 kcal/mol and for birds ranged from -24.7 to -30.2 kcal/mol indicating a lower binding specificity for non-avian species. These results indicate that while ALDH1A3 is found in all amniotes, the targeting of ALDH1A3 could have a unique role in the evolution of the avian integumentary structures.

We found that 50 Tenascin C (TNC) sequences share a similar target site with the chicken. These 50 sequences represent 44 species (39 birds, 3 turtles and 2 crocodilians) indicating that the target site is conserved in Archelosauria (Crawford et al., 2015). TNC is a macromolecule with multiple fibronectin type III repeats and in the chicken there are at least three splice variants (Chiquet-Ehrismann, 1990; Erickson and Bourdon, 1989) and our microarray probe was in a region conserved in all three variants. The largest TNC (MW: 230KD) isoform has 8 fibronectin type III repeats. The other two TNC isoforms lack two (MW: 200KD) and three fibronectin type III repeats (MW: 190KD), respectively, in the chicken (Tucker et al., 1994). Interestingly, the miRNA target site is in fibronectin type III repeat 6 which is only found in the two largest TNC isoforms (Figure 3.3b) suggesting that the targeting of TNC may relate to alternative mRNA

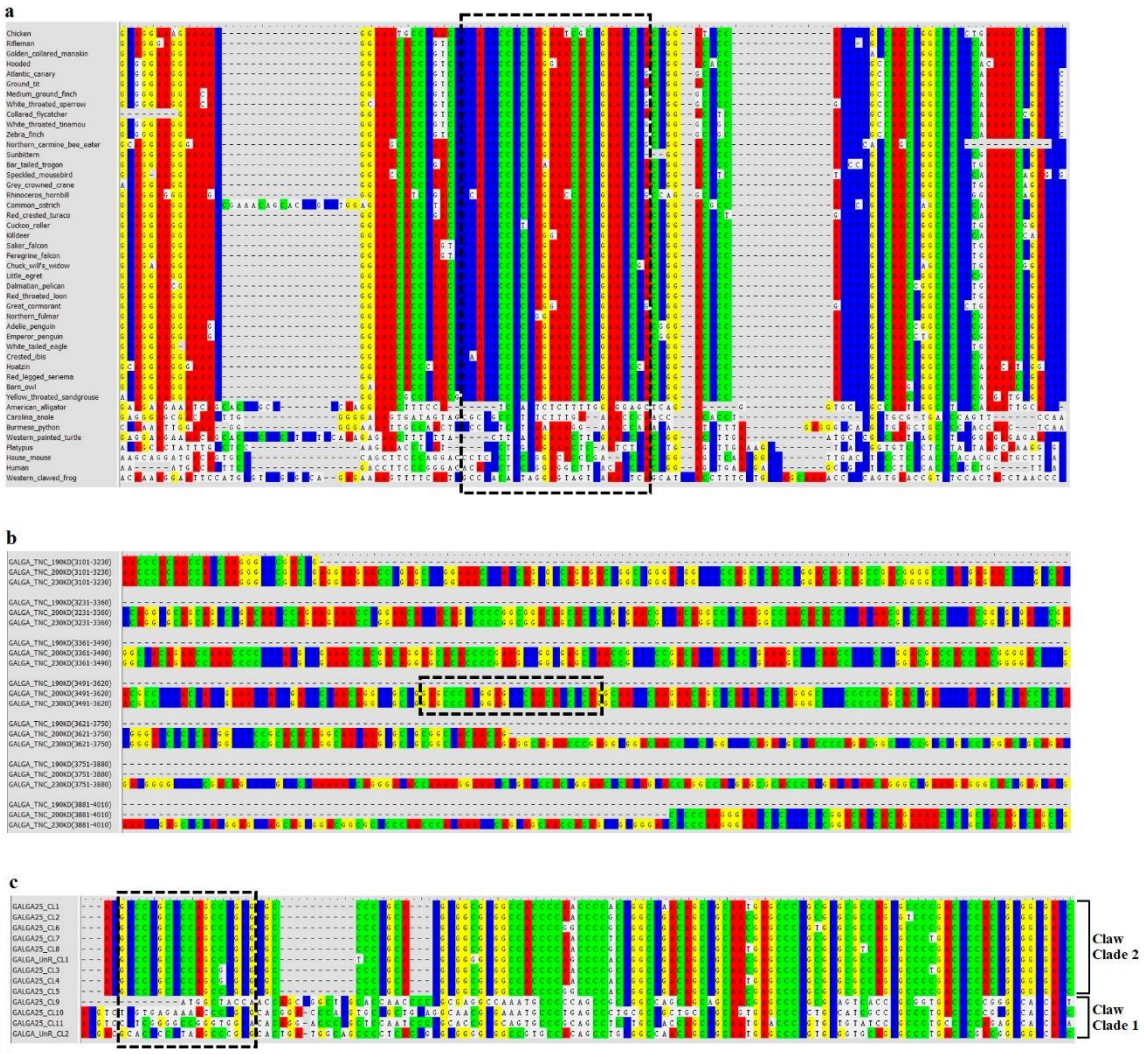


Figure 3.3. (a) Illustrates the nucleotide alignment of Aldhla3 sequences from 45 tetrapod species of which 37 are avian species. The dashed black rectangle highlights the miRNA target site conserved in birds to the exclusion of other tetrapods. (b) Illustrates the nucleotide alignment of three chicken Tenascin C (TNC) alternative splice variants. The dashed black rectangle highlights the miRNA target site that is conserved in two of the three splice variants. (c) Illustrates the nucleotide alignment of 13 chicken claw β-keratins from the two claw β-keratin clades (Greenwold et al. 2014; Zhang et al. 2014). The dashed black rectangle highlights the miRNA target site conserved in the clade 2 claw β-keratins.

splicing. Furthermore, three other birds (Peking duck, pigeon and White-throated

tinamou) also have the TNC splice variants without the conserved region. Interestingly,

TNC is identified as one of the 62 feather associated genes in genomic proximity to the Conserved non-exonic elements (CNEEs) which were identified as possible regulators of closed linked genes (Lowe et al., 2015).

Six claw β-keratins in the DF17/DF19 comparison group (Table 3.1) are all predicted to be targeted by the same miRNA (gga-miR-1556). The predicted target sites of these 6 differentially expressed claw β-keratins are 100% identical. We found that all six of these β-keratins belong to clade 2 claw β-keratins of the β-keratin phylogeny presented in Zhang et al. (2014). Furthermore, a sequence alignment of all of the chicken claw β-keratins demonstrate that the target site is found in the 5' region of the gene and is absent in clade 1 claw β-keratins (Figure 3.3c). Additionally, a search of the 63 avian clade 2 claw β-keratins of the 48 bird genomes (Greenwold et al., 2014) found that 21 clade 2 claw β-keratins contain exactly the same predicted target site while no BLAST hits for this region were found for clade 2 claw β-keratins for 3 species (Chimney swift, Peregrine falcon and Kea). In contrast to the claw β-keratins, the two keratinocyte β-keratins (GALGA25_Ktn11 and GALGA25_Ktn8), while both being members of clade 1 keratinocyte β-keratins (Greenwold et al., 2014; Zhang et al., 2014), are targeted by different miRNAs for different target sites and are the only keratinocyte β¬-keratins possessing their respective target sites. The only feather β-keratin in Table 3.1 is found on chromosome unknown but it is grouped with chromosome 10 feather β-keratins of the chicken (Greenwold et al., 2014; Zhang et al., 2014). These data indicate that sequence and clade specific miRNA targeting of β-keratins are occurring during feather and scutate scale development.

The remaining genes in Table 3.1 have varying target site and miRNA conservation results. While Gastrulation Brain Homeobox 2 (GBX2) has no other miRNA target site BLAST hits outside of the chicken, the miRNA (gga-miR-1759-3p) appears to originate in birds. The two α-keratin genes' (GALGA_5AKI_KRT19 and GALGA_5AKII_KRT5) miRNA target sites were only found in these chicken genes. The targeting of GALGA_5AKII_KRT5 is, however, unique in that the miRNA and target site appears to have evolved in the ancestor of modern chickens. Fibroblast Growth Factor 20 (FGF20) has 24 BLAST hits including 3 birds and multiple mammalian species, but the miRNA appears to have originated in birds. The miRNA target site of Ectodysplasin A receptor (EDAR) is found on many species of tetrapods, but its miRNA is found only in the chicken. Of the 187 solute carrier family 45 member 2 (SLC45A2) genes downloaded from NCBI, 45 BLAST hits are birds and only 3 avian genes from 3 species (Egretta garzetta, Pelecanus crispus and Nipponia Nippon) do not have the predicted conserved region. Only 11 Cadherin 1, type 1 (CDH1) genes found on NCBI had the chicken miRNA target site and all of these genes were bird genes. Furthermore, the miRNA gene, gga-miR-3538, that targets CDH1 is only found in birds. Similar to CDH1is the gene of Ephrin type-A receptor 4 (EPHA4) which only has two target site hits from the chicken and ground tit (Table 3.1; Additional File 3.6).

DISCUSSION

miRNA target prediction

It is difficult to computationally predict the mRNA target sites for miRNAs in organisms because of the limited number of validated miRNA targets and the uncertainty of miRNA binding specificity. Some computational algorithms such as miRanda, TargetScanS and

PicTar are mainly used to search potential miRNA target sites in only the 3' UTR of mRNAs (Krek et al., 2005; Lewis et al., 2005; Pfeffer et al., 2004). However, many recent studies show that miRNAs can also target the coding and 5' untranslated regions of mRNAs (Duursma et al., 2008; Lytle et al., 2007). In this study, we did not restrict the miRNA target site to the 3' UTR. Using the RNAhybrid algorithm, we expanded the target site by searching the entire mRNA sequence of all chicken mRNAs. The RNAhybrid algorithm produces output based on the stability of thermodynamics and also considers the nucleotide base pair complementary binding. Many popular miRNA target prediction programs are only based on the nucleotide base pair complementary binding and do not consider the differential regulation of mRNAs. Therefore, here, we cross compared the RNAhybrid results with the differentially expressed mRNA genes from our microarray data. We also reinforced our RNAhybrid results using RNA22 which is a Markov Chain based algorithm that allows users to adjust the sensitivity and specificity and search the whole mRNA sequence.

Evolution of chicken miRNA genes across tetrapods

Wheeler et al. (2009) analyzed 18 phylogenetically diverse metazoan taxa using Roche 454 sequencing data and genomic searches and found that miRNA families are rarely lost in the descendant lineages after a miRNA emerges. We also found that most of the chicken miRNA genes present in the most recent common ancestor (MRCA) can be found in extant taxa. For example, 79.6% (78 out of the 98) of the chicken miRNA genes found in the frog (Xenopus Silurana tropicalis) can also be found in mammals, squamates, turtles, crocodilians and birds. However, this doesn't imply that miRNA genes are not lost in the descendant lineages such as some of the 48 birds may have

lineage specific gene loss based on our results in Additional File 3.2. The red numbers to the right of the taxa names in Figure 3.2 are the number of miRNA genes that cannot be found in that taxa but are found in the hierarchical most recent common ancestor. For example, 14 miRNA genes in the frog cannot be found in 3 mammals (Ornithorhynchus anatinus, Mus musculus and Homo sapiens) suggesting possible miRNA gene lost in the common ancestor of mammals.

Signaling pathways and miRNAs involved in epidermal development

The chicken scaleless low-line mutant (sc/sc) fails to form both feathers and scales (Abbott and Asmundson, 1957) while the high-line chicken mutant (sc/+) only lack scutate scales and develops feathers on the metatarsal skin where scutate scales are normally located (Sawyer et al., 2005). Fibroblast growth factor 2 (FGF2) was found to rescue the low-line phenotype in that feathers formed upon application of FGF2 to the dorsal skin (Song et al., 1996). More recently, Wells et al. (2012) found that a missense mutation in the third exon of the Fibroblast Growth Factor 20 (FGF20) gene is completely associated with the scaleless low-line chicken mutant. In this study, we identified that FGF20 is likely being targeted by gga-miR-6651-5p in the scale 8 vs. scale 17 and 19 comparison groups. Interestingly, studies using scaleless chickens have demonstrated that scutate scale formation may be contingent on the inhibition of the formation of feather placodes and that this occurs around day 8 of embryogenesis (Dhouailly et al., 1980; Sawyer and Knapp, 2003; Tanaka et al., 1987). In fact, normal feather development begins earlier than scale development (Lucas and Stettenheim, 1972; Rawles, 1963; Tanaka and Kato, 1983). Furthermore, we found that FGF20 is only upregulated in scutate scale day 8 tissue relative to days 17 or 19 which is in agreement

with the results of Wells et al. (2012) which found that FGF20 is expressed in the dorsal epidermis during day 8 of embryogenesis. Thus, we propose that FGF20 expression is important for the initiation of feather placode development and that the miRNA targeting of FGF20 regulates FGF20 expression by repressing feather placode development in metatarsal epidermis around day 8 of embryogenesis and subsequently initializing scale development.

Retinoic acid (RA) synthesizing pathway is important to avian epidermal development (Chuong et al., 1992; Dhouailly et al., 1980; Fisher et al., 1988). Application of RA in chicken during early embryogenesis (~day 7-10) was shown to induce feather fated skin explants to form scale like structures and reticulate scale epithelium to form feather like structures that express feather specific genes (Chuong et al., 1992; Dhouailly et al., 1980; Fisher et al., 1988). Retinoic acid is synthesized from retinol (vitamin A) by two oxidative steps using alcohol dehydrogenases to produce retinaldehyde which is then oxidized into retinoic acid by retinaldehyde dehydrogenases (RALDH1, 2 and 3; now referred to as ALDH1A1, ALDH1A2, and ALDH1A3; Niederreither et al., 2002). In this study we found evidence that ALDH1A3 was targeted by a miRNA at an Aves specific target site and that ALDH1A3 is upregulated in scutate scale day 8 in the SC8/DF8 comparison group and this expression pattern also correlated with the finding from Mou et al. (2011) that AlDH1A3 is distinctively expressed in the chicken hindlimb region during day 7-8 of embryogenesis. A recent study of the Naked neck chicken mutant found that ALDH1A3 inhibits feather placode development (Mou et al., 2011) thereby suggesting that the miRNA regulation of ALDH1A3 is important in

determining the fate of the different regions of the avian epidermis by controlling the level and timing of RA.

Targeting Tenascin C (TNC)

Tenascin C (TNC) is a macromolecule consisting of disulfide-linked subunits and the subunits are composed of structural domains of epidermal growth factor-like repeats and fibronectin type III repeats that act as a cell adhesion molecule (Tucker et al., 1994). Shames et al. (1994) found that TNC was exclusively expressed at Day 18 by preforming in situ hybridization experiment and our microarray transcriptome data is correlated with this expression pattern. The effects of TNC on cell adhesion varies from cell type to cell type in that it can promote cell adhesion or non-adhesiveness or even cell repulsion in different cell types (Bourdon and Ruoslahti, 1989; Chiquet-Ehrismann et al., 1988; Erickson and Taylor, 1987; Joshi et al., 1993; Prieto et al., 1992; Spring et al., 1989). Fischer et al. (1997) demonstrated different levels of adhesion for different TNC variants in that TN-ADC (MW: 230KD) had a higher level of cell adhesion activity than TN-190 (MW: 190KD) for certain cell lines. Studies have found that TNC is expressed during feather and scutate scale development (Fischer et al., 1997; Shames et al. 1994; Tucker, 1991). The role of TNC in the dynamic process of cell adhesion and non-adhesion in feather development is likely a complex and dynamic process where intercellular adhesion is critical for the development of certain parts of the feather while a complete loss of cell adhesion allows for the "branching" of different parts of the feather such as barbs and barbules (see Maderson et al., 2009).

Our results indicate that miRNA gga-miR-146a-5p may play a role in the alternative splicing of TNC by targeting a fibronectin type III repeat missing in only the

190 KD TNC splice variant (Figure 3.3b). In scutate scales of the chicken, TNC isoforms appear in a temporally and spatially restricted manner during development (Shames et al., 1991). While the larger isoforms of TNC (215, 230 and 250 KD) appear in the scutate scale by day 15 of embryogenesis, the 190 KD isoform did not appear until day 18 (Shames et al., 1994). Furthermore, Shames et al. (1994) found that at day 21 of embryogenesis, the 230 and 250 KD isoforms have disappeared from the scutate scale region. Our differential expression results of the TNC miRNA-mRNA duplex along with those of Shames et al. (1994) indicates that gga-miR-146a-5p is likely targeting the larger TNC isoform for degradation starting at day 17 of embryogenesis in the scutate scale thus leaving only the 190 KD TNC isoform at hatching.

It has been experimentally validated that the human miRNA miR-335 can suppress metastasis and cancer cell migration by targeting TNC which can promote the cancer via signaling pathways (Oskarsson et al., 2011; Png et al., 2011). MicroRNA miR-335 is only found in mammals based on the available miRNA database. As previously stated, multiple miRNAs can target the same mRNA and a miRNA can target multiple mRNAs. Our results indicate that TNC is targeted by miR-146a-5p and that miR-146a-5p is found in all amniotes. However, the miRNA target site is only found in TNC genes of birds, crocodilians and turtles, but not mammals indicating that miR-146a-5p may have different gene targets in different organisms. Additionally, the BLAST searches for the TNC target site found hits for all of the crocodilian and turtle TNC mRNA genes including the Chinese alligator which has 4 splice variants possibly suggesting that only birds have a TNC splice variant lacking the miR-146a-5p target site. Together, the data

presented here suggest that miR-146a-5p targeting of TNC may play an exclusive and important role in avian epidermal differentiation.

Targeting β-keratins

Beta (β)-keratins are filamentous (fibrous) proteins (Filshie and Rogers, 1962; Fraser and Parry, 2011) found only in reptiles and birds and are found in beaks, claws, spurs, scales and feathers (Baden and Maderson, 1970; Fraser and Parry, 1996; Gregg and Rogers, 1986; O'Guin and Sawyer, 1982; Sawyer et al., 1986). Recent phylogenetic analyses of avian β-keratins group them into four subfamilies (claw, feather, scale and keratinocyte; Greenwold and Sawyer, 2013) and 6 phylogenetic clades composed of two keratinocyte β-keratin clades, one feather-β-keratin clade, one scale β-keratin clade and two claw β-keratin clades (Greenwold et al., 2014; Zhang et al., 2014). Annotations of each β-keratin gene are based on the species 5 letter abbreviation, chromosome location, subfamily and chromosomal orientation (see detail annotation in Greenwold et al., 2014).

A shotgun proteomic analysis of adult cornified chicken epidermal tissues (beak, claw, scutate scale and feather) identified that the scutate scale had between 0 to 9 unique protein species similar to 3 β-keratins (beta-keratin, claw keratin, scale keratin) (Rice et al., 2013) and two-dimensional gel studies found that only 7 different β-keratin protein species were found in the extracts of chicken scutate scale at day 17 of embryogenesis (Knapp et al., 1993; Shames et al., 1988). Linking this with recent expression data which found that 101 different β-keratin genes were upregulated in embryonic day 17 scutate scale vs day 8 scutate scale (Greenwold et al., 2014) indicates that that a large number of β-keratin mRNAs may be degraded (possibly by miRNAs) prior to mRNA translation. In this study we found that a single miRNA is likely targeting 6 claw β-keratins from the

same phylogenetic clade (Greenwold et al., 2014), two different miRNAs are targeting two different keratinocyte β-keratins and a miRNA is targeting a feather β-keratin (Table 3.1). While this provides compelling evidence that certain β¬-keratin mRNAs are being targeted for degradation, we have likely inadvertently excluded many chicken miRNAs and consequently potential miRNA/β-keratin mRNA duplexes, from our initial microarray experiment. In fact, the number of identified chicken miRNAs on miRBase has grown from 639 to 737 during the period of this study. Furthermore, we only analyzed two β-keratin containing tissues (scutate scale and feather), thus suggesting that miRNAs may play a major role in the structural formation of avian epidermal appendages.

Feather β-keratins form the largest subfamily of β-keratins and evolved in the lineage giving rise to modern day birds (Greenwold et al., 2014; Greenwold and Sawyer, 2010). Interestingly, they are the only avian β-keratins found on multiple chromosomes of the chicken and zebra finch (Greenwold and Sawyer, 2010; Greenwold et al., 2014; Ng et al., 2014). In the chicken, they are found on chromosomes 1, 2, 6, 7, 10, 25 and 27 (Greenwold et al., 2014; Ng et al., 2014). The sole feather β-keratin, GALGA_UnR_FK10, we found to be a likely candidate for miRNA targeting is differentially expressed in the WF17/WF19 comparison group (Table 3.1) and found on chromosome unknown but groups with chromosome 10 feather β-keratins of the chicken (Greenwold et al., 2014; Zhang et al., 2014). There is a growing amount of evidence that feather β-keratins on different chromosomes have different expression profiles. For example, the feather β-keratin found on chromosome 7, barbule specific keratin 1 (BlSK1), was found to be expressed in feather follicles which specifically produce

pennaceous barbules, but no expression was found in the follicles destined to form plumulaceous barbules (Kowata et al., 2014). Furthermore, Greenwold et al. (2014) found different profiles of expression between feather β-keratins on microchromosome 25 and 27 while Ng et al. (2014) found expression profile differences between single feather β-keratins of chromosome 2, 6, 7, 25 and 27 for regenerating contour and flight feathers using in situ hybridization. Together, these data indicate that feather β-keratins on different genomic loci are differentially expressed during feather development and that post-transcriptional modification (miRNAs) of β-keratin mRNAs contribute to the expression patterns at the temporal level along with the regional specific level. These regional patterns of β-keratin mRNA expression may also contribute to the amazing diversity of feather barbule morphology seen among extant birds (Maderson et al., 2009).

CHAPTER 4

USING SCALE AND FEATHER TRAITS FOR MODULE CONSTRUCTION PROVIDES A FUNCTIONAL APPROACH TO CHICKEN EPIDERMAL DEVELOPMENT

Bao, W., Greenwold, M. J., Sawyer, R. H. (2017). *Functional & Integrative Genomics*, 1-11

ABSTRACT

Gene co-expression network analysis has been a research method widely used in systematically exploring gene function and interaction. Using the Weighted Gene Co-expression Network Analysis (WGCNA) approach to construct a gene co-expression network using data from a customized 44K microarray transcriptome of chicken epidermal embryogenesis, we have identified 2 distinct modules that are highly correlated with scale or feather development traits. Signaling pathways related to feather development were enriched in the traditional KEGG pathway analysis and functional terms relating specifically to embryonic epidermal development were also enriched in the Gene Ontology analysis. Significant enrichment annotations were discovered from customized enrichment tools such as Modular Single-Set Enrichment Test (MSET) and Medical Subject Headings (MeSH). Hub genes in both trait-correlated modules showed strong specific functional enrichment toward epidermal development. Also, regulatory elements, such as transcription factors and miRNAs were targeted in the significant enrichment result. This work highlights the advantage of this methodology for functional prediction of genes not previously associated with scale and feather trait related modules.

INTRODUCTION

Avian feathers and scutate scales are intricate structures of birds and perform multiple functions such as flight, thermal regulation and protection (Gill, 1995). Hundreds of genes (i.e. signaling pathway genes, structural genes, cell adhesion genes, etc.) have been associated with the morphogenesis of these two structures (Harris et al. 2002; Widelitz et.al., 1999; Andl et al. 2002; Bell and Thathachari, 1963; Baden and Maderson, 1970; Haake et al. 1984; Rogers, 1985; Fischer et al. 1997, Shames et al. 1994; Tucker, 1991;

Maderson et al. 2009), and gene co-expression network analysis is a new approach to discover gene interactions in a large gene expression datasets (Van Noort et al. 2003; Stuart et al. 2003; Horvath and Dong, 2008; Ng et al. 2014; Chang et al. 2015). It is based on the quantitative pairwise relationships among thousands of gene transcript profiles (Zhou et al. 2002; Steffen et al. 2002; Stuart et al. 2003; Zhang and Horvath, 2005; Carey et al. 2005; Schaefer and Strimmer, 2005; Chuang et al. 2008; Langfelder and Horvath, 2008). Gene network construction has been applied to identify gene pathways and the genes underlying complex traits (Barabasi and Oltvai, 2004; Featherstone and Broadie, 2002; Thieffry et al. 1998).

Here we performed a Weighted Gene Co-expression Network Analysis (WGCNA) using a customized 44K chicken microarray dataset of total RNA extracted from feather and scutate scale tissue during chicken embryonic development (Li et al. 2008). A WGCNA network is defined as sets of cluster nodes with each node representing a single transcript (Langfelder and Horvath, 2008). The edges, which are the connections between nodes, are calculated by the Pearson's correlation coefficient based on the expression profile of the transcripts (Langfelder and Horvath, 2008). WGCNA can identify gene modules which are the clusters of highly interconnected genes (Langfelder and Horvath, 2008) associated with a trait. This method has been successfully applied in the diverse field of genomic research (Xue et al. 2013; Wang et al. 2014; Hicker et al. 2015). There are two types of co-expression networks: condition free and condition dependent networks. The condition free co-expression network normally utilizes large datasets from different conditions including tissue types, biological traits and experimental conditions to construct the basis for genetic architecture of gene expression

profiles. The traditional (condition free or dependent) gene co-expression network analysis is limited by the small group of differential gene expression data which are chosen using arbitrary thresholds. In order to study the specificity of epidermal development in the chicken embryo, we defined and applied the trait conditions (condition dependent network analysis) during epidermal embryogenesis and built modules showing the trait relatedness with all annotated genes in the customized microarray.

MATERIALS AND METHODS

Tissue Samples and microarray experiment

Chicken dorsal feather (DF) and scutate scale (SC) tissues were taken at day 8, 17 and 19 of chicken embryonic development, while wing feather (WF) tissue was taken at day 17 and 19. A customized version of the chicken 44K Agilent microarray was used to analyze mRNA extracted from these tissue samples with a total of 31 individual microarray expression samples analyzed. The details of experiment can be found in Bao et al. (2016) and Greenwold et al. (2014).

Data Set Pre-processing

The raw gene expression Feature Extraction (FE) files produced by Agilent High-Resolution Microarray Scanner (Agilent Technologies; Palo Alto, CA) were imported into Agilent GeneSpring software (13.0 GX version: Agilent Technologies; Palo Alto, CA). The customized probes were 639 miRNA genes, 27 α-keratin genes, 102 β-keratins and 2 Histidine-Rich Protein (HRP) genes. All the probes in the microarray were normalized with the percentile shift algorithm internally and the percentile target was set as the default 75% and the preprocess baseline for all samples was set to the median

option. All the probes with the same Entrez Gene ID and Gene Symbol ID were grouped on average as the same gene and this also applied to the probes having the same customized annotation.

Network construction and module detection

After the normalization and grouping steps of all probe expressions, there were 12,601 genes with Entrez Gene IDs or customized annotations, while 13,118 genes did not have any type of Primary Accession annotations. We used the hclust function in WGCNA to achieve standard hierarchical clustering for the identification of outlier microarray samples and there is no obvious outlier sample (Additional File 4.1) (Langfelder et al. 2008). The clustering result also complies with previous clustering results (Bao et al.

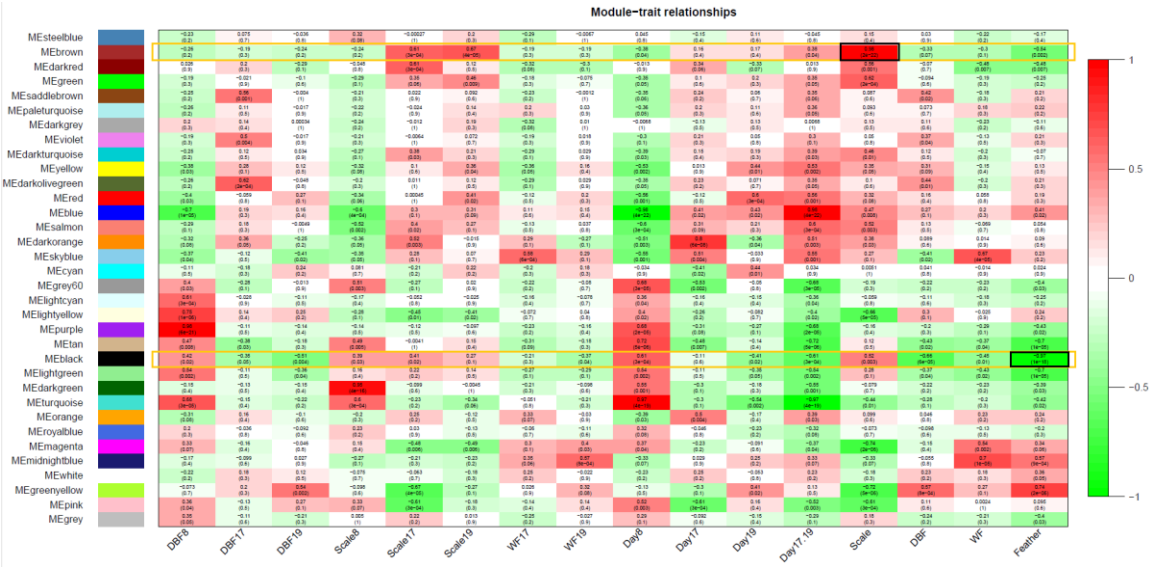


Figure 4.1. Displays the heatmap of Module-trait relationships for the co-expression network. Each row, identified by different colors, represents modules of highly correlated genes and each column corresponds to a defined external trait. Each cell contains the corresponding correlation and p-value for each respective module and trait. The brown module and black module are highlighted with a gold rectangle. The module-scale trait correlation is 0.95 for the brown module and the module-feather trait correlation is -0.97 for the black module.

2016). We define the biological traits (scale, feather, day, stages etc.) as a binary system (1 or 0 represent positive or negative; Additional File 4.2) to relate each detected module. The adjacency of the unsigned network matrix was calculated based on the absolute Pearson correlation coefficient raised to the power β. Soft-thresholding power 12 was determined by the optimum scale free topology which target the balance between scale-free fit index and the mean connectivity (Additional File 4.3). The modules were detected by the calculation of Topological Overlap Measure (TOM) and the static tree was cut by merging height as 0.25 with the minimum module size as 30 genes. All modules were named by a different color (i.e, steelblue, brown, darkred etc.; see also Figure 4.1).

Hub gene detection

The hub genes are defined as the ones having multiple internode connections based on the weight calculation (Langfelder et al. 2008). The intra-modular hub genes within the particular trait related module are more important than the whole network hub genes (Zhang and Horvath, 2005; Horvath et al. 2006; Langfelder et al. 2013). The weight threshold was set at default value (0.05) for the MeSH enrichment analysis of module hub genes.

Traditional Functional annotation and analysis of module genes

KOBAS (KEGG Orthology Based Annotation System, V3.0; http://kobas.cbi.pku.edu.cn/; Xie et al. 2011) was performed to identify biological pathways and Gene Ontology termsfor the genes in 11 modules having both moderate Module-Trait and Module Membership correlations (Additional File 4.4; see below in Results). The foreground genes are the ones having Entrez gene ID annotations and the background was set as the whole genome of the chicken. The KEGG pathway database

(Kanehisa and Goto, 2000) were loaded to perform the hypergeometric/Fisher's exact test and the FDR correction method (Benjamini and Hochberg, 1995). GO classification count of all nonduplicate enriched GO annotations were performed by CateGOrizer (Hu et al. 2008).

Gene enrichment analysis using the Modular Single-Set Enrichment Test (MSET)

Enrichment for chicken epidermal development related genes (EDRGs) of the most significant microarray results, based on the ranking of Gene Significance (GS) value, was assessed using the MSET algorithm (Eisinger et al. 2013). GS value is the correlation measurement based on the gene expression profile and the external traits (equation: GSi = |cor (Genei, Trait)|; Langfelder and Horvath, 2008). Therefore, each gene has its own GS value corresponding to the specific trait and the higher the absolute value of the GS, the more important the gene is to the trait (Langfelder and Horvath, 2008). The gene set of interest, relating to chicken epidermal development, was established from a customized gene set having 302 genes as recently reported by Bao et al. (2016). This customized gene set essentially combines the Lowe et al. (2015) curated feather development related genes and the chicken α- and β- keratins annotated by Greenwold et al. (2014). All background genes were ranked by the descending absolute value of gene significance (GS) for feather and scale traits. The GS value of each gene corresponding to a trait was calculated by WGCNA and the details can be found at Additional File 4.5. The number of genes for each MSET randomization step was determined by the number of genes with GS $\geq$ |0.85|. Each analysis used 5000 simulations for the randomization.

Medical Subject Headings (MeSH)-informed enrichment analysis

We downloaded the enrichment tool package meshr, MeSH database package MeSH.db and the chicken specific annotation package MeSH.Gga.eg.db from Bioconductor (https://www.bioconductor.org) (Tsuyuzaki et al. 2015; Morota et al. 2016). We screened all genes having an Entrez Gene ID for the enrichment analysis since the MeSH terms are widely associated with this particular annotation. After loading all three packages, the hypergeometric test was performed on the background genes (all microarray genes with Entrez Gene IDs). We performed MeSH enrichment analyses on selected module hub genes and the selection of the hub genes in each module were based on the weight ($w \geq 0.05$) and the connectivity ($n \geq 2$) threshold calculated by WGCNA. Significantly enriched (p value ≤ 0.05) MeSH terms can be detected as 3categories: A (Anatomy), D (Chemicals and Drugs) and G (Phenomena and Processes)

RESULTS

Network Structure

We used 26,359 expressed genes from the customized 44K chicken microarray to construct a scale free co-expression network using the Weighted Gene Correlation Network Analysis (WGCNA) algorithm. We identified 33 modules which included most of the genes in the customized microarray. There are 338 genes categorized as non-module genes that are not co-expressed with other module genes. Figure 4.1 is the heatmap of the Module-Trait relationship for the co-expression network. Each row is identified by different colors, representing modules of highly correlated genes and each column corresponds to a defined external trait. Each cell contains the corresponding eigengene (E) significance which is the correlation between each module eigengene and the defined external trait (scale, feather etc.) of the samples. In this study, the eigengene

is the virtual gene of the module that represents the expression profile in the whole module. We found that 9 out of the 33 modules are moderately correlated ($r \geq 0.5$, $p \leq 0.01$) with scale traits (Figure 4.1). Six of the 33 modules are moderately correlated ($r \geq 0.5$, $p \leq 0.01$) with the feather traits (Figure 4.1). It is notable that the brown module ($r = 0.98$, $p = 2e\text{-}22$) and the black module ($r = -0.97$, $p = 1e\text{-}18$) have very strong correlation values with scale and feather traits, respectively.

In contrast to the eigengene significance, each gene in the whole network has its own Gene Significance (GS) which is a measure of the correlation between a single gene and the external traits. In other words, the higher the absolute value of the GS, the more important the gene is to the trait. To further investigate the relationship between the traits and the genes, we plotted the correlation of module membership (MM) and the gene significance (GS) for each module (Figure 4.2). Since the module membership is defined as the correlation between each gene and the modular eigengene, it is reasonable that the modules having high correlation of module membership and gene significance with the external trait tend to be more related to the trait because the eigengene is a more reliable representative gene for the whole module. Based on this intra-modular analysis, we further identified 7 modules with the scale trait and 5 modules with the feather trait that have both moderate correlations ($r > 0.5$, $p < 0.01$) for Module-Trait correlation and MM-GS correlation (Additional File 4.4). One of these modules (greenyellow) correlates with both the scale and feather traits. As Figure 4.2 shows, the brown module with the scale trait and the black module with the feather trait have strong correlations ($r > 0.9$, $p < 0.01$) in the scatter of the MM-GS correlation plot and they are the only modules that also have strong Module-Trait correlations ($r > 0.9$, $p < 0.01$). Therefore, these two modules

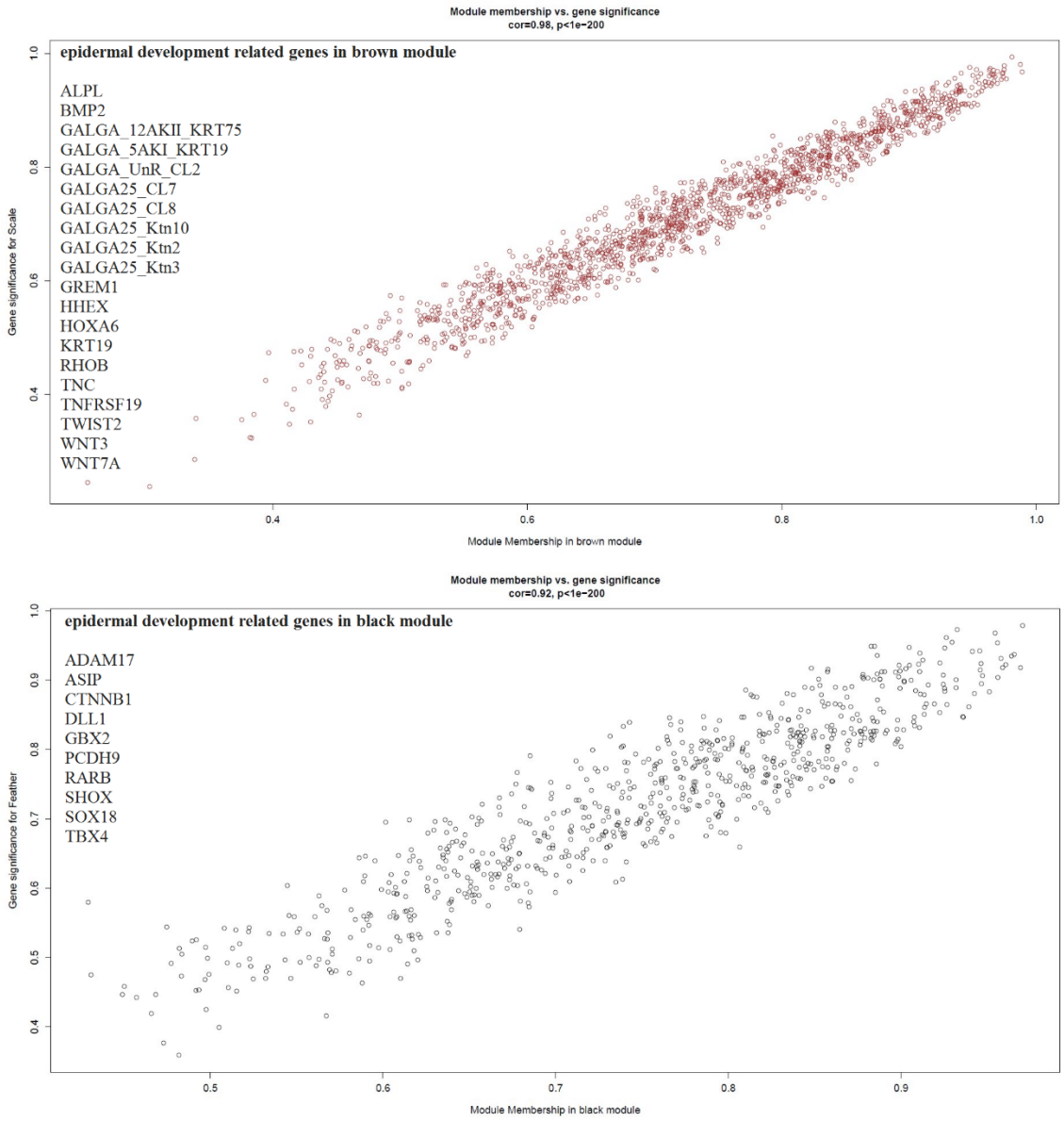


Figure 4.2. Displays the gene scatter plot of correlation for Module Membership (MM) and Gene significance (GS). Each spot represents a single gene. The horizontal axis is the module membership value which represent the connectivity of a single gene and the virtual module representative gene (Eigengene). The vertical axis is the gene significance (GS) value which represents the correlation of a single gene and the external trait. Figure 4.1 shows that brown module has a high module-scale trait correlation and black module also has a high module-feather trait correlation. The top figure shows that the brown module has high correlation of MM-scale GS (r=0.98, p= 2e-22), while the bottom figure shows that the black module (r=-0.97, p=1e-18) has high correlation of MM-feather GS. All the EDRG genes in each module.

have great potential to influence epidermal development during chicken embryogenesis. Below we will present the analyses of these two important modules (brown and black) using novel enrichment methods (MSET and MeST).to understand the chicken epidermal development.

KEGG pathway and Gene Ontology enrichment

We used an integrative functional enrichment tool (KOBAS 3.0) on the 11 unique modules having both moderately correlated Module-Trait and MM-GS relationships to identify the biological pathways and Gene Ontology (GO) terms. We recorded all the significantly enriched KEGG pathways, GO annotations, corrected p-values, and correlated traits for each of the 11 modules (Additional File 4.6). All 11 modules have enriched KEGG pathways and GO annotations. It is notable that some well-known signaling pathways (Wnt and TGF-beta) known to be involved in feather skin differentiation (Chang et al. 2004; Jiang et al. 2004; Chuong et al. 1991) were enriched. There are 1,582 unique enriched GO Ontology terms across the 11 modules and 1,123 (71%) of them are classified as Biological Process root terms.

Enrichment for chicken epidermal development genes using MSET analysis

A set of genes relating to chicken epidermal development (epidermal development related genes; EDRGs) was compiled as stated in the Material and Methods section. We ran 4 sets of MSET analyses using the background genes: all microarray genes using their feather GS value, all microarray genes using their scale GS value, brown module genes using their scale GS value and black module genes using their feather GS value. For the MSET analysis using the whole microarray background (feather and scale GS values), we preprocessed the gene module dataset and obtained 13,241 genes having either Gene

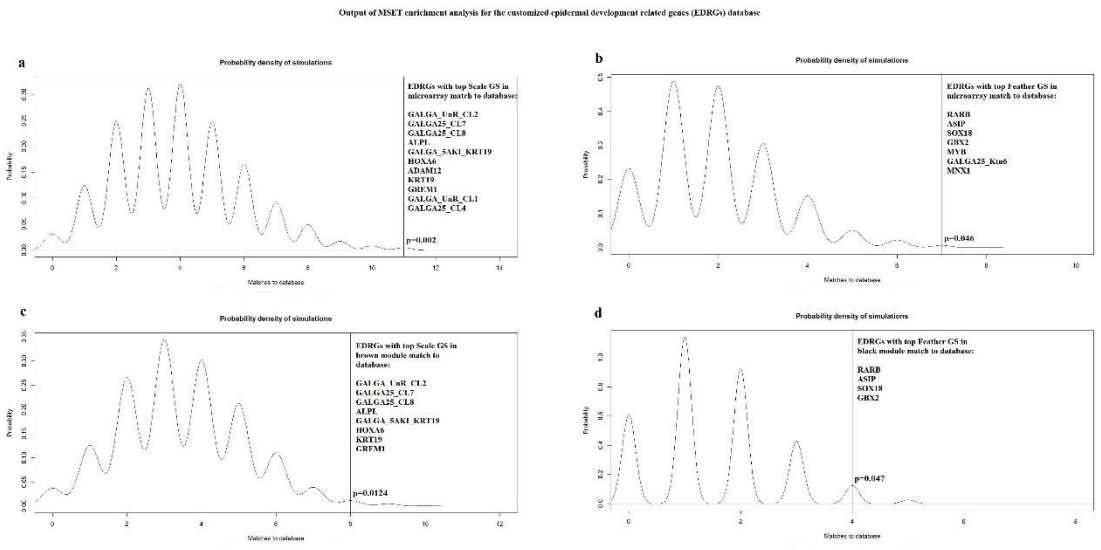


Figure 4.3. Displays the output of MSET enrichment analysis for the customized chicken epidermal development related gene database within the chicken embryonic epidermal development expression data. The background genes for window a. and b. are the whole microarray genes and the importance ranking are based on scale gene significance (GS) and feather gene significance (GS) for each gene, respectively. Likewise, the background genes for window c. and d. are the black module and brown module respectively. The gene importance order of black and brown modules is based on the value of the feather gene significance (GS) and scale gene significance (GS) for each gene. The vertical axis represents the probability of X number of matches to the customized database found in a randomly generated set of simulated results from the background.

Symbol annotations or our customized α- and β-keratin annotations. The enrichment analyses of scale and feather traits are based on the rankings of the Gene Significance (GS) values of scale and feather traits obtained with WGCNA. MSET analyses found significant enrichment in both top (genes with GS $\geq |0.85|$) microarray results (scale and feather traits) indicating that chicken epidermal development genes appear in the significant microarray data at a rate higher than what would be due to random chance (Figure 4.3). Alpha (α-) and beta (β-) keratins were enriched in both of the top list of genes in the analysis based on ranked feather/scale GS value as the background. By

applying the same threshold for the top gene selection under the MSET algorithm, we also performed a MSET analysis on the background genes of the brown and black modules ranked by scale GS (brown module) and feather GS (black module) for each module. Both of these analyses found significant enrichment for the "top genes" based on the trait Gene Significance (GS) threshold (genes with GS ≥ |0.85|; see Material and Method section) and α- and β-keratins are still enriched in the brown module with scale GS (Figure 4.3).

Hub gene enrichment analysis for the highly correlated trait modules

The MeSH database provides 16 different categories of MeSH terms (Tsuyuzaki et al. 2015) derived from the National Library of Medicine (NLM). We chose 3 major categories (A: Anatomy; D: Chemicals and Drugs and G: Phenomena and Processes) as they related directly to this study and ignored other unrelated categories such as "Psychiatry and Psychology", "Geographical Locations", "Humanities" etc. Table 4.1 lists the statistically significant MeSH terms enriched for the hub genes in the brown and black modules. The MeSH term "Chicken Embryo" in the Anatomy (A) category (MeSH: D002642) was enriched in the brown module and this term has the highest number of unique curated literature sources (4836) relating to the background genes of the brown module. Another MeSH term "Gene Expression Regulation, Developmental" (MeSH: D018507) was also significantly enriched in both the brown and black modules. These results depict the scenario that the module hub genes are involved in gene regulation of chicken embryogenic development. The "Feathers" MeSH term (MeSH: D005241) of the Anatomy (A) category was enriched in the black (feather trait) module hub genes while it is not enriched in brown (scale trait) module genes. These findings

Table 4.1. Shows a list all MeSH terms enriched for the hub genes in brown and black modules.

Module with hub genes	**MeSH Category**	**Enriched MeSH Terms**
Black	D: Chemicals and Drugs	Actinin; Fibroblast Growth Factor 10; Receptors, Fibroblast Growth Factor; Multienzyme Complexes; Receptors, Cytoplasmic and Nuclear; Microfilament Proteins; Muscle Proteins; MicroRNAs;
	G: Phenomena and Processes	Wnt Signaling Pathway; Muscle Contraction; DNA Breaks, Double-Stranded; Structure-Activity Relationship; Nucleic Acid Hybridization; Molecular Weight; Gene Expression Regulation, Developmental
	A: Anatomy; D: Chemicals	Interneurons; Actin Cytoskeleton; Hindlimb; Adipose Tissue; Wing; Ear, Inner; Gizzard; Lymphoid Tissue; Muscle, Smooth; Feathers
Brown	D: Chemicals and Drugs	Receptors, Somatostatin; Octoxynol; Neuropilin-1; Neuropilin-2; TNF Receptor-Associated Factor 6; Cystine; Tissue Inhibitor of Metalloproteinase-2; RNA, Small Nucleolar; Matrix Metalloproteinase 9; haker Superfamily of Potassium Channels; Inhibitor of Differentiation Protein 2; Inhibitor of Differentiation Protein 1; Smad5 Protein; eIF-2 Kinase; Sp3 Transcription Factor; terleukin-1 Receptor-Associated Kinases; MAP Kinase Kinase Kinases; Thyroxine; Sp1 Transcription Factor; Proto-Oncogene Proteins c-jun; Reactive Oxygen Species; Toll-Like Receptor 3; Proto-Oncogene Proteins c-maf; Semaphorins; Semaphorin-3A; Antigens, Differentiation; Thyrotropin; Collagen Type I; Ubiquitins; Platelet-Derived Growth Factor; Peptide Library; Alanine; Proto-Oncogene Proteins c-fos; Triiodothyronine; Frizzled Receptors; Viral Proteins; Oligonucleotides, Antisense; Ribosomal Proteins; Receptors, G-Protein-Coupled; Protein Precursors; RNA, Untranslated; Amino Acids; Helix-Loop-Helix Transcription Factors; Single-Chain Antibodies;

		Receptors, Cell Surface; Green Fluorescent Proteins; Homeodomain Proteins; Nerve Tissue Proteins
	G: Phenomena and Processes	Sequence Tagged Sites; Molecular Dynamics Simulation; Gene Order; Chromosomes, Human, Pair 7; Peptide Library; Spermatogenesis; Enzyme Induction; Genome, Human; Osteogenesis; Gene Rearrangement; MAP Kinase Signaling System; Brain Chemistry; Lymphocyte Activation; Enzyme Activation; Down-Regulation; Up-Regulation; Introns; Open Reading Frames; Species Specificity; Gene Library; Signal Transduction; Gene Expression Regulation, Developmental; Sequence Homology, Amino Acid; Amino Acid Sequence
	A: Anatomy; D: Chemicals	Arteries; Tectum Mesencephali; Peripheral Nerves; Synaptic Vesicles; Mandible; Primitive Streak; Chromosomes, Human, Pair 7; Dendrites; resynaptic Terminals; Head; Pericardium; Myocytes, Cardiac Heart Valves; Neurons, Afferent; Lung; Lymphoid Tissue; Eye; Limb Buds; Heart; Extremities; Mesoderm; Neurons; Brain; Chick Embryo;

The enrichment analysis in Table 4.1 was performed on three major categories (A: Anatomy; D: Chemicals and Drugs and G: Phenomena and Processes)

validate the WGCNA module detection approach based on trait correlation and the effectiveness of our strategy indicating that the black module hub genes specifically contribute to feather structure during chicken embryogenesis. It is notable that the

"MicroRNAs" term (MeSH: D035683) was found to be significant and has the most (1,700) unique curated literature sources relating to the background genes in the Chemicals and Drugs (D) category for the black module hub genes.

We selected all edges (gene connections) with nodes (genes) having Gene Symbol annotations, then collected all unique nodes with the frequency calculation. There are 785 and 400 unique genes having at least 2 edges for the brown and black modules, respectively. The highest number of edges for the hub genes in the brown module is 653, while the highest number of edges for the hub genes in the black module is 328. We also found that there are 20 and 10 EDRGs in the brown and black modules, respectively (Additional File 4.7). We ranked these EDRGs based on the number of edges (edge connectivity) in the modules. For each module, more than half of the EDRGs are among the top 50% of all nodes with the most connectivity in each module. Furthermore, three claw β-keratin EDRGs in the brown module and the retinoic acid receptor beta (RARB) gene in the black module ranked among the top 5% of edge connectivity among all the other nodes. Most of the EDRGs in brown or black modules have at least 20 connections and only one EDRG in each module has the minimum 2 connections. Therefore, EDRGs in these highly correlated epidermal traits related modules have high connectivity and likely play an important role in chicken epidermal embryogenesis (Additional File 4.7)

DISCUSSION

In this study, we did not employ the traditional method of using differentially expressed genes to construct the co-expression network. Typically, researchers will screen their datasets to remove genes that are not differentially expressed or lack annotation (Miller et al. 2014; Maschietto et al. 2015; Wong et al. 2015). In contrast, Hudson et al. (2009) used

their entire dataset and found that modifications of transcription factors (TF) such as reversible phosphorylation can act independently of the TF expression levels. Similarly, we took advantage of this method and used a large number of microarray datasets (31 arrays) and we were able to include most of the available expression data on the 44K microarray chips and build a more robust co-expression network related to epidermal developmental traits. The detection of co-expressed modules related to phenotypic traits can facilitate the identification of grouped genes and their regulatory mechanisms. The workflow of this study is summarized in Figure 4.4.

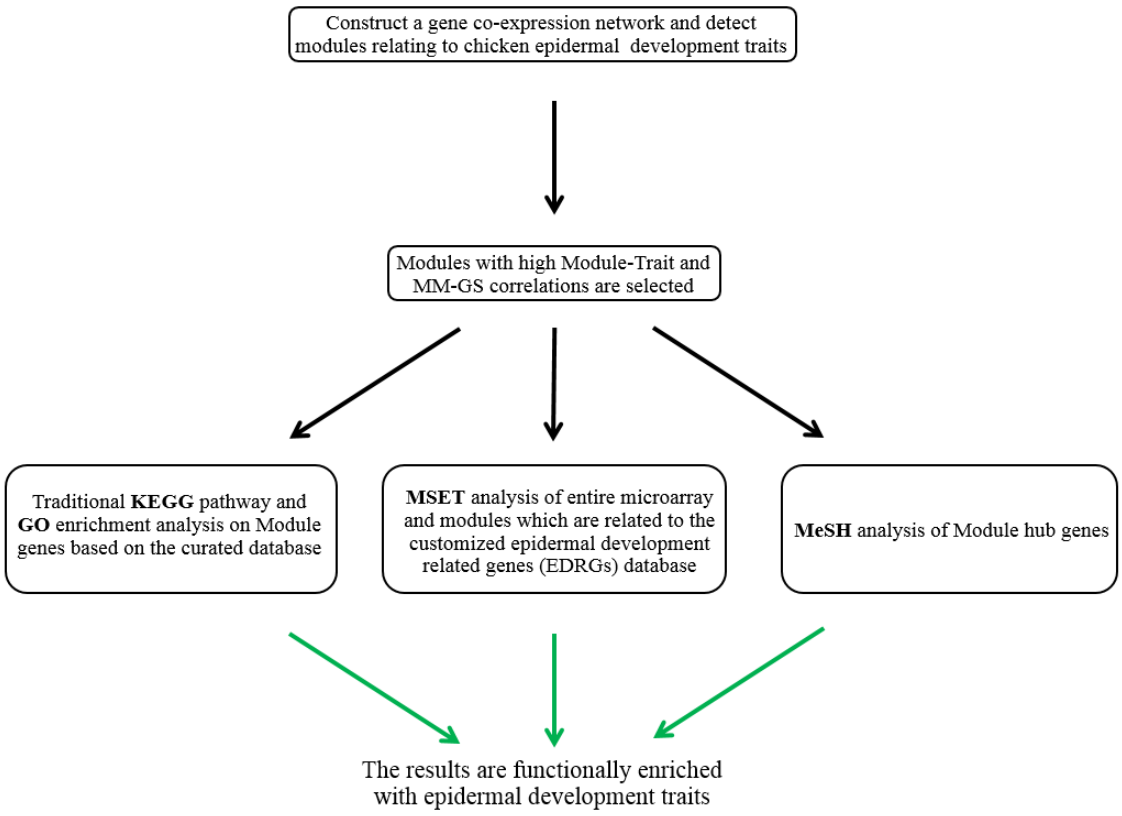


Figure 4.4. illustrates the workflow of this study.

Unlike quantitative traits such as body weight, length etc., the qualitative traits are difficult to measure and to build a Module-Trait relationship. Based on previous work

defining the embryonic stages of the chicken (Hamburger and Hamilton, 1951), we used the strategy of 1 for positive and 0 for negative of Böhne et al. (2014) to discover the modules related to the scale or feather traits. Figure 4.1 not only shows the correlation of the scale and feather traits to the clustered module genes, it also shows correlation with other non-quantitative traits, such as stages (Day 8, Day 17 and Day 19), which will be studied in the future.

Although 11 modules meet the basic correlation threshold ($r > 0.5$) for Module-Trait and MM-GS relationships, only the brown module for the scale trait and the black module for the feather trait have very strong correlations ($r > 0.9$). Therefore, we focused on analyzing these two EDRG modules, which are highly correlated to avian epidermal development. Chang et al. (2015) applied a cosine similarity algorithm to identify the differentially expressed /coregulated genes in chicken developing epidermal tissue and found 5 significantly enriched KEGG pathways: ECM-receptor interaction, Focal adhesion, Melanogenesis, Calcium signaling pathway and Vascular smooth muscle contraction. Their further verification work found that voltage-gated calcium channel subunits were expressed spatio-temporally in the epithelium and this strongly suggests that the calcium signaling pathway is involved in early skin development as a novel pathway. Besides the ECM-receptor interaction pathway, all the other 4 KEGG pathways were also enriched in our analysis. Our first KEGG pathway analysis, using an earlier version of KOBAS (2.0), failed to identify any enriched pathways for the black module because of the missing Entrez Gene ID annotations. This may also explain the missing ECM-receptor interaction pathway enrichment. Although the enriched GO terms are numerous and complicated, we are able to identify informative terms such “skin

development", "embryo development ending in birth or egg hatching" and "morphogenesis of embryonic epithelium" that legitimizes tissue type in our analysis. Interestingly, we found that Tenascin C (TNC) is involved in an enriched Gene Ontology term (GO:0070161). Previous studies have shown that TNC is dynamically expressed and alters the cell adhesiveness during feather and scutate scale development (Fischer et al. 1997; Shames et al. 1994; Tucker, 1991). Furthermore, TNC was also discovered to be a possible miRNA target gene during epidermal development in the chicken embryo (Bao et al. 2016).

The traditional KEGG pathway and Gene Ontology analysis is based on the established and curated pathways and gene database. However, there are no miRNAs or beta-keratins involved in any chicken KEGG pathways. Considering our WGCNA modules contain a large number of these genes, we did not find it appropriate to only perform the KEGG pathway and Gene Ontology analyses and we therefore additionally utilized performed MSET analysis. MSET employs a simple randomization test (Eisinger et al. 2013) allowing comparison of the microarray results with independently curated gene sets that are not commonly associated or categorized in other enrichment tools. In other words, the background genes for the enrichment are customized based on the research interests. In our case, the customized gene set is the EDRGs. It is interesting that the brown module for the scale trait has 5 β-keratins enriched as the top microarray genes. Considering β-keratins are the major structural genes for avian scales and the MSET analysis is sensitive to the gene significance result, this indicates the trait specific modules heavily utilize this specific type of β-keratin.

Topologically, central genes in the co-expression network, the hub genes, have more internode connections. It is reasonable that hub genes are more likely to be required for network integrity and play more important regulatory functions. Therefore, the identification and characterization of hub genes in the network is helpful to understanding the basis of the epidermal development network. However, we cannot simply input the hub genes of a module to perform the KEGG pathway enrichment analysis as the results may not be fruitful because the lack of other non-hub genes associated with pathways may heavily influence the enrichment results. Therefore, we ran the KEGG pathway analysis using KOBAS 3.0 on all the hub genes in black module and brown module and compared it to the analysis with all of the module genes. The enrichment results for all genes in black and brown modules produced a total of 19 enriched pathways while there was no enriched pathway for the hub genes in black module and only one "Ribosome" KEGG pathway enriched for the hub genes in the brown module. The traditional enrichment methods (KEGG pathway and GO enrichment) were not useful in identifying the functional role of modular hub genes.

In contrast to the enriched MeSH Anatomy term "Feathers" (D005241) for the black module with the feather trait, we did not find the "Scale" MeSH term enriched in the brown module based on the scale trait, because there is no anatomical or epidermal term "Scale" deposited in the National Library of Medicine (NLM). However, the brown module hub genes may still be important for the formation of the scale, but are not detectable. Datasets such as the brown module hub genes may be candidates to be deposited in the NLM to improve its number of MeSH terms.

Besides the “Feathers” term enriched in the black module, there are hub genes in the black and brown modules that are enriched for the MeSH Anatomy (A) category, which are related to epidermal development, such as “Wing” (D014921), “Hindlimb” (D006614), and “Limb Buds” (D018878). During embryonic day 9 to 16, the feather buds undergo a complicated process to eventually become the feather follicles, which involves the connection to nerve fibers and muscles to provide physiological functions (Yu et al. 2002; Yu et al. 2004). Interestingly, we found that many muscle or nerve related MeSH terms were enriched either for the Chemicals and Drugs (D) category or Phenomena and Processes (G) category such as “Muscle Contraction” (D009119) and “Nerve Tissue Proteins” (D009419). In addition, we found that the MeSH term “Thyroxine” (D012269) is enriched with a high p-value in the brown module D category. Thyroxine initiates epidermal thickening, secondary periderm formation and even contributes to the cornification of embryonic chicken skin in culture (Tammi and Maibach, 1987). The classic “Wnt Signaling Pathway” (D060449) is enriched as the top ranked G term in the black module, and some other enriched MeSH terms such as ‘Signal Transduction” (D015398) indicates that the signaling transduction pathways are involved, but the data are too general for annotation.

We noticed that there are 2 genes of the SMADs protein family present in our customized feather related gene set, and that the SMAD5 protein was another gene family member enriched as MeSH D category in the brown module. Ng et al. (2015) conducted a transcriptome analysis of regenerating adult feathers of the chicken and found that SMAD5 was significantly enriched as one of the differentially expressed genes in multiple functional enrichment annotations. Considering our MeSH enrichment results,

we suggest that SMAD5 is involved in signaling pathway regulation as early as the embryonic developmental stages examined here. Lowe et al. (2015) curated and collected the feather development related genes with specific standards such as feather phenotype mutation or defined region of the feather, thus it is inevitable that they missed some genes contributing to epidermal development, such as those found by Ng et al. (2015) in their transcriptome analysis. Therefore, our MeSH enrichment results based on module hub genes may provide a new approach for selection of candidate genes involved in avian epidermal development.

We also found that 2 transcriptional factors (TF) genes (specific protein 1 [Sp1], specific protein 3 [Sp3]) are enriched by MeSH. Ng et al (2015) found that Sp3 was significantly upregulated in the late body feather development. Sp1 was predicted by IPA (Ingenuity Pathway Analysis) canonical pathway analysis to be an upstream regulator gene contributing to the gene expression differences between distal and proximal body feathers in the chicken. These results together with other pathway enriched terms suggest that module hub genes play important roles in TF regulation to coordinate with the biological signaling cascade process of chicken epidermal embryogenesis. This research has utilized the co-expression network and various enrichment strategies to characterize the most highly featured gene modules instead of exhaustively listing every single ontology. The entire gene regulatory network is very complicated when we consider the thousands of target gene possibilities produced by TF and miRNA genes. Hopefully, this study will shed light on our understanding of the co-expression networks and the important role of hub genes in the featured modules.

CONCLUSIONS

The detection of chicken co-expression network modules during embryonic epidermal development in the chicken is useful for understanding the relatedness between modules and the traits via customized enrichment methods. Specifically, for the modules with strong correlations of high Module-Trait and MM-GS values, the MSET enrichment results demonstrate the important functional role of epidermal related genes in both whole genome and modular datasets. The novel application of MeSH enrichment for identifying the hub genes, which have more gene connections, allows for the discovery of their functional roles in the formation of chicken epidermal structures.

CHAPTER 5

CONCLUSIONS

The epidermal appendages especially feathers are diverse in birds. Generations of scientists have been studying the development and evolution of these complex structures for years. To understand the evolution of the diversification of the feathers, I first utilized the 48 bird genomes to study the two important structural genes (α- and β- keratins), which exist in the epidermal appendages of reptiles and birds (Chapter 2). This includes identifying the chromosomal orientation of each gene for all avian species, building phylogenetic trees for the gene families and comparing the subfamily members of the structural genes. I found that the number of α-keratin genes is reduced while the β-keratin multi-gene family has undergone great expansions in the avian lineage. I also proposed that there has been dynamic evolution of the β-keratins and each subfamily, which has allowed birds to occupy habitats world-wide. The comparison of β-keratin copy numbers and the discovery that the relative proportions for each subfamily in different birds with diverse lifestyles (aquatic, terrestrial, predatory) differ, supports the important role of β-keratins in avian ecology. The dynamic expansion or contraction of the keratin multi-gene families have provided the structural variation needed. In Chapter 2, I also performed the comparative transcriptome analyses for α- and β-keratins from three developmentally important regions of the epidermis (scale region, dorsal and wing feather regions) in the chicken at various embryonic stages. The expression of feather β-keratins on different genomic loci for dorsal feather and wing feather in the Day 17 embryos suggests that the evolution of diverse feather β-keratins at different loci contributed to the formation of epidermal structures with new physical properties.

In contrast to the expression data, only a few beta keratin proteins can be isolated from epidermal tissues (scale, feather) at the same developmental stages by 2-dimentional

gel electrophoresis. This discrepancy suggests that β-keratin mRNAs may be inactivated possibly by miRNAs. Hence profiling the miRNAs during chicken epidermal embryogenesis and finding their potential target genes are important steps to understanding how miRNAs exert their functions during scale and feather development.

In Chapter 3, I utilized the transcriptome data and identified miRNA-mRNA duplexes using a highly cited miRNA prediction tool with stringent threshold and filtering of the results. MicroRNA regulation networks can be very complex considering that each miRNA may target multiple mRNAs at multiple sites. I endeavored to identify a manageable number of miRNA-mRNA duplexes from thousands of possibilities. Moreover, all of the protein coding genes I identified in Table 3.1 of Chapter 3 have been previously studied and reported as correlating with avian epidermal differentiation, so I summarized the most relevant genes into 3 categories (cell signaling genes, cell adhesion and structural genes) and highlighted the representative gene in each category. Furthermore, there are no prior studies that report on the mechanism of the expression between discrepancy mRNA and protein expression for the β-keratins, which are the most important structural genes for avian epidermal appendages. Chapter 3 demonstrates that multiple β-keratins are targeted by miRNAs, which provides a mechanism to explain the expression data and also suggests an important role of miRNAs in contributing to the formation of structurally diverse avian epidermal appendages.

In Chapter 4, I first applied the WGCNA method to construct modules related to epidermal development traits. The results revealed 11 unique modules with high Module-Trait and Module Membership-Gene Significance (GS) correlations. The traditional KEGG pathway and Gene Ontology enrichment analyses on these modules demonstrated

the involvement of classic signaling pathways. To study the functional role of structural genes (α- and β-keratins) and miRNAs during chicken embryonic development, two modules with extremely high Module-Trait and MM-GS correlations were selected to perform a novel MSET analysis based on a customized epidermal development database. I also use a new MeSH method to investigate the functional role of hub genes in trait related modules. The results of both MSET and MeSH are functionally enriched with epidermal development traits.

Recent studies have shown that the early morphogenesis of scales and feather use similar signaling gene network pathways; such as beta-catenin, sonic hedgehog, Bmp2, and the wnt gene family (Harris et al., 2002; Musser et al., 2015). These studies conclude that feathers are derived from earlier developmental stages of an asymmetric scale present in the archosaurian ancestor. While these networks are undoubtedly important to feather and scale morphogenesis numerous other regulatory genes and their networks are involved. In fact, Lowe et al. (2014) found that 86% of the regulatory elements and 100% of the nonkeratin feather gene set were present prior to the origin of dinosaurs. While previous studies compare similar morphological stages for scale and feather development (the epidermal placode) they do not consider the actual days of development, which are very different. The epidermal placodes of alligator scales appear after 20 days of development; the epidermal placodes of chicken scutate scales occurs at 9.5 days of development, and the feather epidermal placode occurs at 6.5 days of development. Clearly many different regulatory gene networks are being expressed at these very different stages of development. For example, the spatial patterns of epidermal cell proliferation differ greatly for the feather and scale epidermal placodes (Sawyer, 1972a,

b; Sawyer et al. 2003; 2005) suggesting a role for regulatory genes such as let-7 and lin-4, which are involved in developmental timing (Skipper, 2008; Martinez et al., 2008).

Chang et al. (2015) have examined the differentially regulated pathways associated with developing feather and scale regions of the chicken skin. Again, they use different days of development when comparing the feather and scale tissues. They find that 5 significant KEGG regulatory pathways are differentially regulated between feather and scale, with a 5-fold increase in the calcium signaling pathway in the 9-day scale epidermis as compared to the 7-day feather epidermis. They find similar divergence in other several genes involved in the calcium signaling pathway between the feather and scale tissue epidermis and dermis. This suggests that at the placode stage, the epidermis and dermis of scales and feathers are using a number of differentially expressed genes.

In my analyses of gene co-expression networks, comparing scale and feather tissues at the same stage of development, I found that the scale and feather epidermal tissues use different epidermal development related genes (EDRGs) suggesting that feathers did not evolve from an ancestral asymmetrical scale, but evolved directly from the ancestral archosaurian epidermis, which had already evolved the novel embryonic cell population known as the subperiderm.

REFERENCES

1000 Genomes Project Consortium. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature*, 491(7422), 56-65.

Abascal, F., Zardoya, R., & Posada, D. (2005). ProtTest: selection of best-fit models of protein evolution. *Bioinformatics*, 21(9), 2104-2105.

Abbott, U. K., & Asmundson, V. S. (1957). Scaleless, an inherited ectodermal defect in the domestic fowl. *Journal of Heredity*, 48(2), 63-70.

Alföldi, J., Di Palma, F., Grabherr, M., Williams, C., Kong, L., Mauceli, E., ... & Ray, D. A. (2011). The genome of the green anole lizard and a comparative analysis with birds and mammals. *Nature*, 477(7366), 587-591.

Alibardi, L. (2013). Immunolocalization of alpha-keratins and feather beta-proteins in feather cells and comparison with the general process of cornification in the skin of mammals. *Annals of Anatomy-Anatomischer Anzeiger*, 195(2), 189-198.

Alibardi, L., & Sawyer, R. H. (2002). Immunocytochemical analysis of beta (β) keratins in the epidermis of chelonians, lepidosaurians, and archosaurians. *Journal of Experimental Zoology Part A: Ecological Genetics and Physiology*, 293(1), 27-38.

Alibardi, L., & Sawyer, R. H. (2006). Cell structure of developing downfeathers in the zebrafinch with emphasis on barb ridge morphogenesis. *Journal of Anatomy*, 208(5), 621-642.

Alibardi, L., & Toni, M. (2008). Cytochemical and molecular characteristics of the process of cornification during feather morphogenesis. *Progress in Histochemistry and Cytochemistry*, 43(1), 1-69.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3), 403-410.

Andl, T., Reddy, S. T., Gaddapara, T., & Millar, S. E. (2002). WNT signals are required for the initiation of hair follicle development. *Developmental Cell*, 2(5), 643-653.

Baden, H. P., & Maderson, P. F. A. (1970). Morphological and biophysical identification of fibrous proteins in the amniote epidermis. *Journal of Experimental Zoology Part A: Ecological Genetics and Physiology*, 174(2), 225-232.

Bao, W., Greenwold, M. J., & Sawyer, R. H. (2016). Expressed miRNAs target feather related mRNAs involved in cell signaling, cell adhesion and structure during chicken epidermal development. *Gene*, 591(2), 393-402.

Barabasi, A. L., & Oltvai, Z. N. (2004). Network biology: understanding the cell's functional organization. *Nature Reviews Genetics*, 5(2), 101-113.

Barnes, G. L. J. (1994). Differentiation of the embryonic skin of the chicken.

Barnes, G. L., & Sawyer, R. H. (1995). Histidine‐rich protein B of embryonic feathers is present in the transient embryonic layers of scutate scales. *Journal of Experimental Zoology Part A: Ecological Genetics and Physiology*, 271(4), 307-314.

Bartel, D. P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, 116(2), 281-297.

Bell, E., & Thathachari, Y. T. (1963). Development of feather keratin during embryogenesis of the chick. *The Journal of Cell Biology*, 16(2), 215-223.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 289-300.

Bereiter-Hahn, J., Matoltsy, A. G., & Richards, K. S. *Biology of the Integument 2 Vertebrates*. 1986.

Betel, D., Wilson, M., Gabow, A., Marks, D. S., & Sander, C. (2008). The microRNA. org resource: targets and expression. *Nucleic Acids Research*, 36(suppl_1), D149-D153.

Birney, E., Clamp, M., & Durbin, R. (2004). GeneWise and genomewise. *Genome Research*, 14(5), 988-995.

Birn-Jeffery, A. V., Miller, C. E., Naish, D., Rayfield, E. J., & Hone, D. W. (2012). Pedal claw curvature in birds, lizards and Mesozoic dinosaurs–complicated categories and compensating for mass-specific and phylogenetic control. *PLoS One*, 7(12), e50555.

Blumenberg, M. (1988). Concerted gene duplications in the two keratin gene families. *Journal of Molecular Evolution*, 27(3), 203-211.

Böhne, A., Sengstag, T., & Salzburger, W. (2014). Comparative transcriptomics in East African cichlids reveals sex-and species-specific expression and new candidates for sex differentiation in fishes. *Genome Biology and Evolution*, 6(9), 2567-2585.

Bourdon, M. A., & Ruoslahti, E. (1989). Tenascin mediates cell attachment through an RGD-dependent receptor. *The Journal of Cell Biology*, 108(3), 1149-1155.

Burt, A., & Bell, G. (1987). Mammalian chiasma frequencies as a test of two theories of recombination. *Nature*, 326(6115), 803-805.

Carey, V. J., Gentry, J., Whalen, E., & Gentleman, R. (2004). Network structures and algorithms in Bioconductor. *Bioinformatics*, 21(1), 135-136.

Carrington, J. C., & Ambros, V. (2003). Role of microRNAs in plant and animal development. *Science*, 301(5631), 336-338.

Carver, W. E., & Sawyer, R. H. (1988). Avian scale development: XI. Immunoelectron microscopic localization of α and β Keratins in the Scutate Scale. *Journal of Morphology*, 195(1), 31-43.

Carver, W. E., & Sawyer, R. H. (1989). Immunocytochemical localization and biochemical analysis of α and β keratins in the avian lingual epithelium. *Developmental Dynamics*, 184(1), 66-75.

Castoe, T. A., De Koning, A. J., Hall, K. T., Card, D. C., Schield, D. R., Fujita, M. K., ... & Reyes-Velasco, J. (2013). The Burmese python genome reveals the molecular basis for extreme adaptation in snakes. *Proceedings of the National Academy of Sciences*, 110(51), 20645-20650.

Chang, C. H., Jiang, T. X., Lin, C. M., Burrus, L. W., Chuong, C. M., & Widelitz, R. (2004). Distinct Wnt members regulate the hierarchical morphogenesis of skin regions (spinal tract) and individual feathers. *Mechanisms of Development*, 121(2), 157-171.

Chang, K. W., Huang, N. A., Liu, I. H., Wang, Y. H., Wu, P., Tseng, Y. T., ... & Oyang, Y. J. (2015). Emergence of differentially regulated pathways associated with the development of regional specificity in chicken skin. *BMC Genomics*, 16(1), 22.

Chen, K., & Rajewsky, N. (2007). The evolution of gene regulation by transcription factors and microRNAs. *Nature Reviews Genetics*, 8(2), 93-103.

Chiquet-Ehrismann, R. U. T. H. (1990). What distinguishes tenascin from fibronectin? *The FASEB Journal*, 4(9), 2598-2604.

Chiquet-Ehrismann, R., Kalla, P., Pearson, C. A., Beck, K., & Chiquet, M. (1988). Tenascin interferes with fibronectin action. *Cell*, 53(3), 383-390.

Chuang, C. L., Jen, C. H., Chen, C. M., & Shieh, G. S. (2008). A pattern recognition approach to infer time-lagged genetic interactions. *Bioinformatics*, 24(9), 1183-1190.

Chuong, C. M., & Homberger, D. G. (2003). Development and evolution of the amniote integument: current landscape and future horizon. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, 298(1), 1-11.

Chuong, C. M., Chen, H. M., Jiang, T. X., & Chia, J. (1991). Adhesion molecules in skin development: morphogenesis of feather and hair. *Annals of the New York Academy of Sciences*, 642(1), 263-280.

Chuong, C. M., Ting, S. A., Widelitz, R. B., & Lee, Y. S. (1992). Mechanism of skin morphogenesis. II. Retinoic acid modulates axis orientation and phenotypes of skin appendages. *Development*, 115(3), 839-852.

Crawford, N. G., Parham, J. F., Sellas, A. B., Faircloth, B. C., Glenn, T. C., Papenfuss, T. J., ... & Simison, W. B. (2015). A phylogenomic analysis of turtles. *Molecular Phylogenetics and Evolution*, 83, 250-257.

Csermely, D., & Rossi, O. (2006). Bird claws and bird of prey talons: Where is the difference? *Italian Journal of Zoology*, 73(01), 43-53.

Dalla Valle, L., Nardi, A., & Alibardi, L. (2010). Isolation of a new class of cysteine-glycine-proline-rich beta‐proteins (beta‐keratins) and their expression in snake epidermis. *Journal of Anatomy*, 216(3), 356-367.

Dalla Valle, L., Nardi, A., Belvedere, P., Toni, M., & Alibardi, L. (2007). β-keratins of differentiating epidermis of snake comprise glycin-proline-serine-rich proteins with an avian-like gene organization. Developmental Dynamics, 236(7), 1939-1953.

Dalla Valle, L., Nardi, A., Bonazza, G., Zuccal, C., Emera, D., & Alibardi, L. (2010). Forty keratin-associated β-proteins (β-keratins) form the hard layers of scales, claws, and adhesive pads in the green anole lizard, Anolis carolinensis. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, 314(1), 11-32.

Dalla Valle, L., Nardi, A., Gelmi, C., Toni, M., Emera, D., & Alibardi, L. (2009a). β-keratins of the crocodilian epidermis: composition, structure, and phylogenetic relationships. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, 312(1), 42-57.

Dalla Valle, L., Nardi, A., Toni, M., Emera, D., & Alibardi, L. (2009b). Beta-keratins of turtle shell are glycine-proline-tyrosine rich proteins similar to those of crocodilians and birds. *Journal of Anatomy*, 214(2), 284-300.

Dhouailly, D., Hardy, M. H., & Sengel, P. (1980). Formation of feathers on chick foot scales: a stagedependent morphogenetic response to retinoic acid. *Development*, 58(1), 63-78.

Duursma, A. M., Kedde, M., Schrier, M., Le Sage, C., & Agami, R. (2008). miR-148 targets human DNMT3b protein coding region. *RNA*, 14(5), 872-877.

Eisinger, B. E., Saul, M. C., Driessen, T. M., & Gammie, S. C. (2013). Development of a versatile enrichment analysis tool reveals associations between the maternal brain and mental health disorders, including autism. *BMC Neuroscience*, 14(1), 147.

Erickson, H. P., & Bourdon, M. A. (1989). Tenascin: an extracellular matrix protein prominent in specialized embryonic tissues and tumors. *Annual Review of Cell Biology*, 5(1), 71-92.

Erickson, H. P., & Taylor, H. C. (1987). Hexabrachion proteins in embryonic chicken tissues and human tumors. *The Journal of Cell Biology*, 105(3), 1387-1394.

Featherstone, D. E., & Broadie, K. (2002). Wrestling with pleiotropy: genomic and topological analysis of the yeast gene expression network. *Bioessays*, 24(3), 267-274.

Filshie, B. K., & Rogers, G. E. (1962). An electron microscope study of the fine structure of feather keratin. *The Journal of Cell Biology*, 13(1), 1-12.

Fischer, D., Tucker, R. P., Chiquet-Ehrismann, R., & Adams, J. C. (1997). Cell-adhesive responses to tenascin-C splice variants involve formation of fascin microspikes. *Molecular Biology of the Cell*, 8(10), 2055-2075.

Fisher, C. J., Knapp, L. W., & Sawyer, R. H. (1988). Retinoic acid induction of featherlike structures from reticulate scales. *Teratology*, 38(4), 321-328.

Fowler, D. W., Freedman, E. A., & Scannella, J. B. (2009). Predatory functional morphology in raptors: interdigital variation in talon size is related to prey restraint and immobilisation technique. *PLoS One*, 4(11), e7999.

Fraser, R. B., & Parry, D. A. (2008). Molecular packing in the feather keratin filament. *Journal of Structural Biology*, 162(1), 1-13.

Fraser, R. B., & Parry, D. A. (2011). The structural basis of the filament-matrix texture in the avian/reptilian group of hard β-keratins. *Journal of Structural Biology*, 173(2), 391-405.

Fraser, R. B., & Parry, D. A. (2014). Amino acid sequence homologies in the hard keratins of birds and reptiles, and their implications for molecular structure and physical properties. *Journal of Structural Biology*, 188(3), 213-224.

Fraser, R. D. B., & Parry, D. A. D. (1996). The molecular structure of reptilian keratin. *International Journal of Biological Macromolecules*, 19(3), 207-211.

Fuchs, E., & Marchuk, D. (1983). Type I and type II keratins have evolved from lower eukaryotes to form the epidermal intermediate filaments in mammalian skin. *Proceedings of the National Academy of Sciences*, 80(19), 5857-5861.

Gardner, P. P., Fasold, M., Burge, S. W., Ninova, M., Hertel, J., Kehr, S., ... & Stadler, P. F. (2015). Conservation and losses of non-coding RNAs in avian genomes. *PLoS One*, 10(3), e0121797.

Garland Jr, T., Dickerman, A. W., Janis, C. M., & Jones, J. A. (1993). Phylogenetic analysis of covariance by computer simulation. *Systematic Biology*, 42(3), 265-292.

Gill, F. B. (1995). Ornithology. Macmillan.

Glenn, T. C., French, J. O., Heincelman, T. J., Jones, K. L., & Sawyer, R. H. (2008). Evolutionary relationships among copies of feather beta (β) keratin genes from several avian orders. *Integrative and Comparative Biology*, 48(4), 463-475.

Greenwold, M. J., & Sawyer, R. H. (2010). Genomic organization and molecular phylogenies of the beta (β) keratin multigene family in the chicken (Gallus gallus) and zebra finch (Taeniopygia guttata): implications for feather evolution. *BMC Evolutionary Biology*, 10(1), 148.

Greenwold, M. J., & Sawyer, R. H. (2011). Linking the molecular evolution of avian beta (β) keratins to the evolution of feathers. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, 316(8), 609-616.

Greenwold, M. J., & Sawyer, R. H. (2013). Molecular evolution and expression of archosaurian β-keratins: Diversification and expansion of archosaurian β-keratins and the origin of feather β-keratins. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, 320(6), 393-405.

Greenwold, M. J., Bao, W., Jarvis, E. D., Hu, H., Li, C., Gilbert, M. T. P., ... & Sawyer, R. H. (2014). Dynamic evolution of the alpha (α) and beta (β) keratins has accompanied integument diversification and the adaptation of birds into novel lifestyles. *BMC Evolutionary Biology*, 14(1), 249.

Gregg, K., & Rogers, G. E. (1986). Feather keratin: composition, structure and biogenesis. In Biology of the integument (pp. 666-694). Springer Berlin Heidelberg.

Griffiths-Jones, S., Saini, H. K., van Dongen, S., & Enright, A. J. (2007). miRBase: tools for microRNA genomics. *Nucleic Acids Research*, 36(suppl_1), D154-D158.

Haake, A. R., König, G., & Sawyer, R. H. (1984). Avian feather development: relationships between morphogenesis and keratinization. *Developmental Biology*, 106(2), 406-413.

Hallahan, D. L., Keiper‐Hrynko, N. M., Shang, T. Q., Ganzke, T. S., Toni, M., Dalla Valle, L., & Alibardi, L. (2009). Analysis of gene expression in gecko digital adhesive pads indicates significant production of cysteine-and glycine-rich beta-keratins. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, 312(1), 58-73.

Hamburger, V., & Hamilton, H. L. (1951). A series of normal stages in the development of the chick embryo. *Journal of Morphology*, 88(1), 49-92.

Harmon, L. J., Weir, J. T., Brock, C. D., Glor, R. E., & Challenger, W. (2007). GEIGER: investigating evolutionary radiations. *Bioinformatics*, 24(1), 129-131.

Harris, M. P., Fallon, J. F., & Prum, R. O. (2002). Shh-Bmp2 signaling module and the evolutionary origin and diversification of feathers. *Journal of Experimental Zoology Part A: Ecological Genetics and Physiology*, 294(2), 160-176.

Hartl, M., & Bister, K. (1995). Specific activation in jun-transformed avian fibroblasts of a gene (bkj) related to the avian beta-keratin gene family. *Proceedings of the National Academy of Sciences*, 92(25), 11731-11735.

Hatzfeld, M., & Franke, W. W. (1985). Pair formation and promiscuity of cytokeratins: formation in vitro of heterotypic complexes and intermediate-sized filaments by homologous and heterologous recombinations of purified polypeptides. *The Journal of Cell Biology*, 101(5), 1826-1841.

He, L., & Hannon, G. J. (2004). MicroRNAs: small RNAs with a big role in gene regulation. *Nature Reviews Genetics*, 5(7), 522-531.

Heller, S., Sheane, C. A., Javed, Z., & Hudspeth, A. J. (1998). Molecular markers for cell types of the inner ear and candidate genes for hearing disorders. *Proceedings of the National Academy of Sciences*, 95(19), 11400-11405.

Hellsten, U., Harland, R. M., Gilchrist, M. J., Hendrix, D., Jurka, J., Kapitonov, V., ... & Blitz, I. L. (2010). The genome of the Western clawed frog Xenopus tropicalis. *Science*, 328(5978), 633-636.

Hesse, M., Zimek, A., Weber, K., & Magin, T. M. (2004). Comprehensive analysis of keratin gene clusters in humans and rodents. *European Journal of Cell Biology*, 83(1), 19-26.

Hickner, P. V., Mori, A., Zeng, E., Tan, J. C., & Severson, D. W. (2015). Whole transcriptome responses among females of the filariasis and arbovirus vector mosquito Culex pipiens implicate TGF-β signaling and chromatin modification as key drivers of diapause induction. *Functional & Integrative Genomics*, 15(4), 439.

Hillier, L. W., Miller, W., Birney, E., Warren, W., Hardison, R. C., Ponting, C. P., ... & Dodgson, J. B. (2004). Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature, 432(7018), 695-716.

Hoeppner, M. P., Gardner, P. P., & Poole, A. M. (2012). Comparative analysis of RNA families reveals distinct repertoires for each domain of life. *PLoS Computational Biology*, 8(11), e1002752.

Horvath, S., & Dong, J. (2008). Geometric interpretation of gene coexpression network analysis. *PLoS Computational Biology*, 4(8), e1000117.

Horvath, S., Zhang, B., Carlson, M., Lu, K. V., Zhu, S., Felciano, R. M., ... & Lee, Y. (2006). Analysis of oncogenic signaling networks in glioblastoma identifies ASPM as a molecular target. *Proceedings of the National Academy of Sciences*, 103(46), 17402-17407.

Hsu, S. D., Tseng, Y. T., Shrestha, S., Lin, Y. L., Khaleel, A., Chou, C. H., ... & Jian, T. Y. (2014). miRTarBase update 2014: an information resource for experimentally validated miRNA-target interactions. *Nucleic Acids Research*, 42(D1), D78-D85.

Hu, Z. L., Bao, J., & Reecy, J. M. (2008). CateGOrizer: A Web-Based Program to Batch Analyze Gene On-tology Classification Categories. *Online Journal of Bioinformatics*, 9(2), 108-112.

Hudson, N. J., Reverter, A., Wang, Y., Greenwood, P. L., & Dalrymple, B. P. (2009). Inferring the transcriptional landscape of bovine skeletal muscle by integrating co-expression networks. *PloS One*, 4(10), e7249.

Huth T. J: The Genomic Organization of the Four Subfamilies of β-keratin (Claw, Feather, Feather-like, and Scale) in the Chicken, *Gallus gallus*. Masters Thesis. 2008, University of South Carolina, Department of Biology.

Jarvis, E. D., Mirarab, S., Aberer, A. J., Li, B., Houde, P., Li, C., ... & Suh, A. (2014). Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science*, 346(6215), 1320-1331.

Jiang, T. X., Widelitz, R. B., Shen, W. M., Will, P., Wu, D. Y., Lin, C. M., ... & Chuong, C. M. (2004). Integument pattern formation involves genetic and epigenetic controls: feather arrays simulated by digital hormone models. *The International Journal of Developmental Biology*, 48, 117.

Joshi, P., Chung, C. Y., Aukhil, I., & Erickson, H. P. (1993). Endothelial cells adhere to the RGD domain and the fibrinogen-like terminal knob of tenascin. *Journal of Cell Science*, 106(1), 389-400.

Kamanu, T. K., Radovanovic, A., Archer, J. A., & Bajic, V. B. (2013). Exploration of miRNA families for hypotheses generation. *Scientific Reports*, 3.

Kane, M. D., Jatkoe, T. A., Stumpf, C. R., Lu, J., Thomas, J. D., & Madore, S. J. (2000). Assessment of the sensitivity and specificity of oligonucleotide (50mer) microarrays. *Nucleic Acids Research*, 28(22), 4552-4557.

Kanehisa, M., & Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, 28(1), 27-30.

Kemp, D. J. (1975). Unique and repetitive sequences in multiple genes for feather keratin. *Nature*, 254(5501), 573-577.

Kemp, D. J., & Rogers, G. E. (1972). Differentiation of avian keratinocytes. Characterization and relations of the keratin proteins of adult and embryonic feathers and scales. *Biochemistry*, 11(6), 969-975.

Kent, W. J. (2002). BLAT—the BLAST-like alignment tool. *Genome Research*, 12(4), 656-664.

Knapp, L. W., Linser, P. J., Carver, W. E., & Sawyer, R. H. (1991). Biochemical identification and immunological localization of two non-keratin polypeptides associated with the terminal differentiation of avian scale epidermis. *Cell and Tissue Research*, 265(3), 535-545.

Knapp, L. W., Shames, R. B., Barnes, G. L., & Sawyer, R. H. (1993). Region‐specific patterns of beta keratin expression during avian skin development. *Developmental Dynamics*, 196(4), 283-290.

Kowata, K., Nakaoka, M., Nishio, K., Fukao, A., Satoh, A., Ogoshi, M., ... & Takeuchi, S. (2014). Identification of a feather β-keratin gene exclusively expressed in pennaceous barbule cells of contour feathers in chicken. *Gene*, 542(1), 23-28.

Krek, A., Grün, D., Poy, M. N., Wolf, R., Rosenberg, L., Epstein, E. J., ... & Rajewsky, N. (2005). Combinatorial microRNA target predictions. *Nature Genetics*, 37(5), 495-500.

Landauer, W., & Dunn, L. C. (1930). The "Frizzle" characters of fowls: Its Expression and Inheritance. *Journal of Heredity*, 21(7), 291-305.

Langfelder, P., & Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, 9(1), 559.

Langfelder, P., Mischel, P. S., & Horvath, S. (2013). When is hub gene selection better than standard meta-analysis?. *PloS One*, 8(4), e61505.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., ... & Thompson, J. D. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*, 23(21), 2947-2948.

Larsson, A. (2014). AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics*, 30(22), 3276-3278.

Lee, L. D., & Baden, H. P. (1976). Organisation of the polypeptide chains in mammalian keratin. *Nature*, 264(5584), 377-379.

Lewis, B. P., Burge, C. B., & Bartel, D. P. (2005). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*, 120(1), 15-20.

Li, X., Chiang, H. I., Zhu, J., Dowd, S. E., & Zhou, H. (2008). Characterization of a newly developed chicken 44K Agilent microarray. *BMC Genomics*, 9(1), 60.

Li, Y. I., Kong, L., Ponting, C. P., & Haerty, W. (2013). Rapid evolution of Beta-keratin genes contribute to phenotypic differences that distinguish turtles and birds from other reptiles. *Genome Biology and Evolution*, 5(5), 923-933.

Lin, C. M., Jiang, T. X., Widelitz, R. B., & Chuong, C. M. (2006). Molecular signaling in feather morphogenesis. *Current Opinion in Cell Biology*, 18(6), 730-741.

Lockhart, D. J., & Winzeler, E. A. (2000). Genomics, gene expression and DNA arrays. *Nature*, 405(6788), 827-836.

Lowe, C. B., Clarke, J. A., Baker, A. J., Haussler, D., & Edwards, S. V. (2014). Feather development genes and associated regulatory innovation predate the origin of Dinosauria. *Molecular Biology and Evolution*, 32(1), 23-28.

Lu, Z. (2011). PubMed and beyond: a survey of web tools for searching biomedical literature. *Database*, 2011.

Lucas, A. M., & Stettenheim, P. R. (1972). *Agriculture handbook*. US Department of Agriculture.

Lytle, J. R., Yario, T. A., & Steitz, J. A. (2007). Target mRNAs are repressed as efficiently by microRNA-binding sites in the 5' UTR as in the 3' UTR. *Proceedings of the National Academy of Sciences*, 104(23), 9667-9672.

Maderson, P. F., Hillenius, W. J., Hiller, U., & Dove, C. C. (2009). Towards a comprehensive model of feather regeneration. *Journal of Morphology*, 270(10), 1166-1208.

Martinez, N. J., Ow, M. C., Barrasa, M. I., Hammell, M., Sequerra, R., Doucette-Stamm, L., ... & Walhout, A. J. (2008). A C. elegans genome-scale microRNA network

contains composite feedback motifs with high flux capacity. Genes & Development, 22(18), 2535-2549.

Maschietto, M., Tahira, A. C., Puga, R., Lima, L., Mariani, D., da Silveira Paulsen, B., ... & Palha, J. A. (2015). Co-expression network of neural-differentiation genes shows specific pattern in schizophrenia. *BMC Medical Genomics*, 8(1), 23.

Miller, J. A., Ding, S. L., Sunkin, S. M., Smith, K. A., Ng, L., Szafer, A., ... & Arnold, J. M. (2014). Transcriptional landscape of the prenatal human brain. *Nature*, 508(7495), 199-206.

Miranda, K. C., Huynh, T., Tay, Y., Ang, Y. S., Tam, W. L., Thomson, A. M., ... & Rigoutsos, I. (2006). A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes. *Cell*, 126(6), 1203-1217.

Miska, E. A., Alvarez-Saavedra, E., Townsend, M., Yoshii, A., Šestan, N., Rakic, P., ... & Horvitz, H. R. (2004). Microarray analysis of microRNA expression in the developing mammalian brain. *Genome Biology*, 5(9), R68.

Mlitz, V., Strasser, B., Jaeger, K., Hermann, M., Ghannadan, M., Buchberger, M., ... & Eckhart, L. (2014). Trichohyalin-like proteins have evolutionarily conserved roles in the morphogenesis of skin appendages. *Journal of Investigative Dermatology*, 134(11), 2685-2692.

Moll, R., Divo, M., & Langbein, L. (2008). The human keratins: biology and pathology. *Histochemistry and Cell Biology*, 129(6), 705-733.

Morota, G., Beissinger, T. M., & Peñagaricano, F. (2016). MeSH-informed enrichment analysis and MeSH-guided semantic similarity among functional terms and gene products in chicken. *G3: Genes, Genomes, Genetics*, 6(8), 2447-2453.

Mou, C., Pitel, F., Gourichon, D., Vignoles, F., Tzika, A., Tato, P., ... & Painter, K. J. (2011). Cryptic patterning of avian skin confers a developmental facility for loss of neck feathering. *PLoS Biology*, 9(3), e1001028.

Musser, J., Gunter, G. P., Prum, R. O. (2013). Investigating the Homology of Feathers and Scales with High-Throughput Genomics. presentation in *Evolution 2013*, Snowbird, Utah, USA

Musser, J. M., Wagner, G. P., & Prum, R. O. (2015). Nuclear β-catenin localization supports homology of feathers, avian scutate scales, and alligator scales in early development. Evolution & Development, 17(3), 185-194.

Ng, C. S., Chen, C. K., Fan, W. L., Wu, P., Wu, S. M., Chen, J. J., ... & Lin, Z. S. (2015). Transcriptomic analyses of regenerating adult feathers in chicken. *BMC Genomics*, 16(1), 756.

Ng, C. S., Wu, P., Fan, W. L., Yan, J., Chen, C. K., Lai, Y. T., ... & Ho, M. R. (2014). Genomic organization, transcriptomic analysis, and functional characterization of avian α-and β-keratins in diverse feather forms. *Genome Biology and Evolution*, 6(9), 2258-2273.

Ng, C. S., Wu, P., Foley, J., Foley, A., McDonald, M. L., Juan, W. T., ... & Leal, S. M. (2012). The chicken frizzle feather is due to an α-keratin (KRT75) mutation that causes a defective rachis. *PLoS Genetics*, 8(7), e1002748.

Niederreither, K., Fraulob, V., Garnier, J. M., Chambon, P., & Dollé, P. (2002). Differential expression of retinoic acid-synthesizing (RALDH) enzymes during fetal development and organ differentiation in the mouse. *Mechanisms of Development*, 110(1), 165-171.

O'Donnell, I. J., & Inglis, A. S. (1974). Amino acid sequence of a feather keratin from Silver Gull (Larus novae-hollandiae) and comparison with one from Emu (Dromaius novae-hollandiae). *Australian Journal of Biological Sciences*, 27(4), 369-382.

O'Guin, W. M., & Sawyer, R. H. (1982). Avian scale development: VIII. Relationships between morphogenetic and biosynthetic differentiation. *Developmental Biology*, 89(2), 485-492.

Oskarsson, T., Acharyya, S., Zhang, X. H., Vanharanta, S., Tavazoie, S. F., Morris, P. G., ... & Massagué, J. (2011). Breast cancer cells produce tenascin C as a metastatic niche component to colonize the lungs. *Nature Medicine*, 17(7), 867-874.

Pabisch, S., Puchegger, S., Kirchner, H. O. K., Weiss, I. M., & Peterlik, H. (2010). Keratin homogeneity in the tail feathers of Pavo cristatus and Pavo cristatus mut. alba. *Journal of Structural Biology*, 172(3), 270-275.

Pfeffer, S., Zavolan, M., Grässer, F. A., Chien, M., Russo, J. J., Ju, J., ... & Tuschl, T. (2004). Identification of virus-encoded microRNAs. *Science*, 304(5671), 734-736.

Png, K. J., Yoshida, M., Zhang, X. H. F., Shu, W., Lee, H., Rimner, A., ... & King, T. A. (2011). MicroRNA-335 inhibits tumor reinitiation and is silenced through genetic and epigenetic mechanisms in human breast cancer. *Genes & Development*, 25(3), 226-231.

Powell, B. C., & Rogers, G. E. (1979). Isolation of messenger RNA coding for the “fast” protein of embryonic chick feathers. *Nucleic Acids Research*, 7(8), 2165-2176.

Presland, R. B., Gregg, K., Molloy, P. L., Morris, C. P., Crocker, L. A., & Rogers, G. E. (1989a). Avian keratin genes I. A molecular analysis of the structure and expression of a group of feather keratin genes. *Journal of Molecular Biology*, 209(4), 549-559.

Presland, R. B., Whitbread, L. A., & Rogers, G. E. (1989b). Avian keratin genes II. Chromosomal arrangement and close linkage of three gene families. *Journal of Molecular Biology*, 209(4), 561-576.

Prieto, A. L., Andersson-Fisone, C., & Crossin, K. L. (1992). Characterization of multiple adhesive and counteradhesive domains in the extracellular matrix protein cytotactin. *The Journal of Cell Biology*, 119(3), 663-678.

Prin, F., & Dhouailly, D. (2003). How and when the regional competence of chick epidermis is established: feathers vs. scutate and reticulate scales, a problem en route to a solution. *International Journal of Developmental Biology*, 48(2-3), 137-148.

Prum, R. O. (1999). Development and evolutionary origin of feathers. *The Journal of experimental zoology*, 285(4), 291-306.

Prum, R. O., & Brush, A. H. (2003). Which came first, the feather or the bird. *Scientific American*, 288(3), 60-69.

Rawles, M. E. (1963). Tissue interactions in scale and feather development as studied in dermal-epidermal recombinations. *Development*, 11(4), 765-789.

Rehmsmeier, M. (2006). Prediction of microRNA targets. *MicroRNA Protocols*, 87-99.

Rehmsmeier, M., Steffen, P., Höchsmann, M., & Giegerich, R. (2004). Fast and effective prediction of microRNA/target duplexes. *RNA*, 10(10), 1507-1517.

Rice, R. H., Winters, B. R., Durbin-Johnson, B. P., & Rocke, D. M. (2013). Chicken corneocyte cross-linked proteome. *Journal of Proteome Research*, 12(2), 771-776.

Rijke, A. M. (1970). Wettability and phylogenetic development of feather structure in water birds. *Journal of Experimental Biology*, 52(2), 469-479.

Rogers, G. E. (1985). Genes for hair and avian keratins. *Annals of the New York Academy of Sciences*, 455(1), 403-425.

Ross-Ibarra, J. (2004). The evolution of recombination under domestication: a test of two hypotheses. *The American Naturalist*, 163(1), 105-112.

Sawyer, R. H. (1972a). Avian scale development. I. Histogenesis and morphogenesis of the epidermis and dermis during formation of the scale ridge. *Journal of Experimental Zoology Part A: Ecological Genetics and Physiology*, 181(3), 365-383.

Sawyer, R. H. (1972b). Avian scale development. II. A study of cell proliferation. *Journal of Experimental Zoology Part A: Ecological Genetics and Physiology*, 181(3), 385-407.

Sawyer, R. H., & Abbott, U. K. (1972). Defective histogenesis and morphogenesis in the anterior shank skin of the scaleless mutant. *Journal of Experimental Zoology Part A: Ecological Genetics and Physiology*, 181(1), 99-110.

Sawyer, R. H., & Fallon, J. F. (1983). Epithelial-mesenchymal interactions in development. *Praeger Publishers*.

Sawyer, R. H., & Knapp, L. W. (2003). Avian skin development and the evolutionary origin of feathers. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, 298(1), 57-72.

Sawyer, R. H., Abbott, U. K., & Fry, G. N. (1974a). Avian scale development. III. Ultrastructure of the keratinizing cells of the outer and inner epidermal surfaces of the scale ridge. *Journal of Experimental Zoology Part A: Ecological Genetics and Physiology*, 190(1), 57-69.

Sawyer, R. H., Abbott, U. K., & Fry, G. N. (1974b). Avian scale development. IV. Ultrastructure of the anterior shank skin of the scaleless mutant. *Journal of Experimental Zoology Part A: Ecological Genetics and Physiology*, 190(1), 71-77.

Sawyer, R. H., Glenn, T. C., French, J. O., & Knapp, L. W. (2005). Developing antibodies to synthetic peptides based on comparative DNA sequencing of multigene families. *Methods in Enzymology*, 395, 636-652.

Sawyer, R. H., Glenn, T., French, J. O., Mays, B., Shames, R. B., Barnes, Jr, G. L., ... & Ishikawa, Y. (2000). The expression of beta (β) keratins in the epidermal appendages of reptiles and birds. *American Zoologist*, 40(4), 530-539.

Sawyer, R. H., Knapp, L. W., & O'Guin, W. M. (1986). Epidermis, dermis and appendages. In *Biology of the Integument* (pp. 194-238). Springer Berlin Heidelberg.

Sawyer, R. H., Rogers, L., Washington, L., Glenn, T. C., & Knapp, L. W. (2005). Evolutionary origin of the feather epidermis. *Developmental Dynamics*, 232(2), 256-267.

Sawyer, R. H., Salvatore, B. A., Potylicki, T. T. F., French, J. O., Glenn, T. C., & Knapp, L. W. (2003). Origin of feathers: Feather beta (β) keratins are expressed in discrete epidermal cell populations of embryonic scutate scales. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, 295(1), 12-24.

Schäfer, J., & Strimmer, K. (2004). An empirical Bayes approach to inferring large-scale gene association networks. *Bioinformatics*, 21(6), 754-764.

Schweizer, J., Bowden, P. E., Coulombe, P. A., Langbein, L., Lane, E. B., Magin, T. M., ... & Wright, M. W. (2006). New consensus nomenclature for mammalian keratins. *The Journal of Cell Biology*, 174(2), 169-174.

Sengel, P. (1976). *Morphogenesis of skin*. Cambridge university press.

Shaffer, H. B., Minx, P., Warren, D. E., Shedlock, A. M., Thomson, R. C., Valenzuela, N., ... & Borchert, G. M. (2013). The western painted turtle genome, a model for the evolution of extreme physiological adaptations in a slowly evolving lineage. *Genome Biology*, 14(3), R28.

Shames, R. B., & Sawyer, R. H. (1986). Expression of β keratin genes during skin development in normal and scsc chick embryos. *Developmental Biology*, 116(1), 15-22.

Shames, R. B., & Sawyer, R. H. (1987). Expression of β-keratin genes during development of avian skin appendages. *Current Topics in Developmental Biology*, 22, 235-253.

Shames, R. B., Bade, B. C., & Sawyer, R. H. (1994). Role of epidermal–dermal tissue interactions in regulating tenascin expression during development of the chick scutate scale. *Journal of Experimental Zoology Part A: Ecological Genetics and Physiology*, 269(4), 349-366.

Shames, R. B., Jennings, A. G., & Sawyer, R. H. (1991). The initial expression and patterned appearance of tenascin in scutate scales is absent from the dermis of the scaleless (sc/sc) chicken. *Developmental Biology*, 147(1), 174-186.

Shames, R. B., Knapp, L. W., Carver, W. E., & Sawyer, R. H. (1988). Identification, expression, and localization of β keratin gene products during development of avian scutate scales. *Differentiation*, 38(2), 115-123.

Shames, R. B., Knapp, L. W., Carver, W. E., Washington, L. D., & Sawyer, R. H. (1989). Keratinization of the outer surface of the avian scutate scale: interrelationship of alpha and beta keratin filaments in a cornifying tissue. *Cell and Tissue Research*, 257(1), 85-92.

Shedlock, A. M., & Edwards, S. V. (2009). Amniotes (amniota). *The timetree of life*, 375-379.

Skipper, M. (2008). A tiny missing link for regulatory networks. Nature Reviews Genetics, 9(11).

Smoak, K. D., & Sawyer, R. H. (1983). Avian spur development: abnormal morphogenesis and Keratinization in the Scaleless (sc/sc) mutant. *Transactions of the American Microscopical Society*, 135-144.

Song, H., Wang, Y., & Goetinck, P. F. (1996). Fibroblast growth factor 2 can replace ectodermal signaling for feather development. *Proceedings of the National Academy of Sciences*, 93(19), 10246-10249.

Spring, J., Beck, K., & Chiquet-Ehrismann, R. (1989). Two contrary functions of tenascin: dissection of the active sites by recombinant tenascin fragments. *Cell*, 59(2), 325-334.

St John, J. A., Braun, E. L., Isberg, S. R., Miles, L. G., Chong, A. Y., Gongora, J., ... & Burgess, S. C. (2012). Sequencing three crocodilian genomes to illuminate the evolution of archosaurs and amniotes. *Genome Biology*, 13(1), 415.

Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, 22(21), 2688-2690.

Steffen, M., Petti, A., Aach, J., D'haeseleer, P., & Church, G. (2002). Automated modelling of signal transduction networks. *BMC Bioinformatics*, 3(1), 34.

Strasser, B., Mlitz, V., Hermann, M., Rice, R. H., Eigenheer, R. A., Alibardi, L., ... & Eckhart, L. (2014). Evolutionary origin and diversification of epidermal barrier proteins in amniotes. *Molecular Biology and Evolution*, 31(12), 3194-3205.

Strasser, B., Mlitz, V., Hermann, M., Tschachler, E., & Eckhart, L. (2015). Convergent evolution of cysteine-rich proteins in feathers and hair. *BMC Evolutionary Biology*, 15(1), 82.

Stuart, J. M., Segal, E., Koller, D., & Kim, S. K. (2003). A gene-coexpression network for global discovery of conserved genetic modules. *Science*, 302(5643), 249-255.

Suzuki, R., & Shimodaira, H. (2006). Pvclust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics*, 22(12), 1540-1542.

Takada, T., Ebata, T., Noguchi, H., Keane, T. M., Adams, D. J., Narita, T., ... & Obata, Y. (2013). The ancestor of extant Japanese fancy mice contributed to the mosaic genomes of classical inbred strains. *Genome Research*, 23(8), 1329-1338.

Tammi, R., & Maibach, H. (1987). Skin organ culture: why?. *International Journal of Dermatology*, 26(3), 150-160.

Tanaka, S., & Kato, Y. (1983). Epigenesis in developing avian scales. I. Qualitative and quantitative characterization of finite cell populations. *Journal of Experimental Zoology Part A: Ecological Genetics and Physiology*, 225(2), 257-269.

Tanaka, S., Sugihara-Yamamoto, H., & Kato, Y. (1987). Epigenesis in developing avian scales: III. Stage-specific alterations of the developmental program caused by 5-bromodeoxyuridine. *Developmental Biology*, 121(2), 467-477.

Thieffry, D., Huerta, A. M., Pérez‐Rueda, E., & Collado-Vides, J. (1998). From specific gene regulation to genomic networks: a global analysis of transcriptional regulation in Escherichia coli. *Bioessays*, 20(5), 433-440.

Tsuyuzaki, K., Morota, G., Ishii, M., Nakazato, T., Miyazaki, S., & Nikaido, I. (2015). MeSH ORA framework: R/Bioconductor packages to support MeSH over-representation analysis. *BMC Bioinformatics*, 16(1), 45.

Tucker, R. P. (1991). The sequential expression of tenascin mRNA in epithelium and mesenchyme during feather morphogenesis. *Development Genes and Evolution*, 200(2), 108-112.

Tucker, R. P., Spring, J., Baumgartner, S., Martin, D., Hagios, C., Poss, P. M., & Chiquet-Ehrismann, R. (1994). Novel tenascin variants with a distinctive pattern of expression in the avian embryo. *Development*, 120(3), 637-647.

Van Noort, V., Snel, B., & Huynen, M. A. (2003). Predicting gene function by conserved co-expression. *Trends in Genetics*, 19(5), 238-242.

Vandebergh, W., & Bossuyt, F. (2012). Radiation and functional diversification of alpha keratins during early vertebrate evolution. *Molecular Biology and Evolution*, 29(3), 995-1004.

Vanhoutteghem, A., Londero, T., Djian, P., & Ghinea, N. (2004). Serial cultivation of chicken keratinocytes, a composite cell type that accumulates lipids and synthesizes a novel β-keratin. *Differentiation*, 72(4), 123-137.

Walker, I. D., & Bridgen, J. (1976). The keratin chains of avian scale tissue. *The FEBS Journal*, 67(1), 283-293.

Walker, I. D., & Rogers, G. E. (1976a). Differentiation in avian keratinocytes. *The FEBS Journal*, 69(2), 329-339.

Walker, I. D., & Rogers, G. E. (1976b). The structural basis for the heterogeneity of chick down feather keratin. *The FEBS Journal*, 69(2), 341-350.

Wang, J., Zhang, X., Shi, M., Gao, L., Niu, X., Te, R., ... & Zhang, W. (2014). Metabolomic analysis of the salt-sensitive mutants reveals changes in amino acid and fatty acid composition important to long-term salt stress in Synechocystis sp. PCC 6803. *Functional & Integrative Genomics*, 14(2), 431.

Wang, Z., Pascual-Anaya, J., Zadissa, A., Li, W., Niimura, Y., Huang, Z., ... & Wang, B. (2013). The draft genomes of soft-shell turtle and green sea turtle yield insights into the development and evolution of the turtle-specific body plan. *Nature Genetics*, 45(6), 701-706.

Warren, W. C., Clayton, D. F., Ellegren, H., Arnold, A. P., Hillier, L. W., Künstner, A., ... & Heger, A. (2010). The genome of a songbird. *Nature*, 464(7289), 757-762.

Warren, W. C., Hillier, L. W., Graves, J. A. M., Birney, E., Ponting, C. P., Grützner, F., ... & Yang, S. P. (2008). Genome analysis of the platypus reveals unique signatures of evolution. *Nature*, 453(7192), 175-183.

Weiss, I. M., & Kirchner, H. O. (2011). Plasticity of two structural proteins: alpha-collagen and beta-keratin. *Journal of the Mechanical Behavior of Biomedical Materials*, 4(5), 733-743.

Wells, K. L., Hadad, Y., Ben-Avraham, D., Hillel, J., Cahaner, A., & Headon, D. J. (2012). Genome-wide SNP scan of pooled DNA reveals nonsense mutation in FGF20 in the scaleless line of featherless chickens. *BMC Genomics*, 13(1), 257.

Wheeler, B. M., Heimberg, A. M., Moy, V. N., Sperling, E. A., Holstein, T. W., Heber, S., & Peterson, K. J. (2009). The deep evolution of metazoan microRNAs. *Evolution & Development*, 11(1), 50-68.

Whitbread, L. A., Gregg, K., & Rogers, G. E. (1991). The structure and expression of a gene encoding chick claw keratin. *Gene*, 101(2), 223-229.

Widelitz, R. B., Jiang, T. X., Chen, C. W., Stott, N. S., & Chuong, C. M. (1999). Wnt-7a in feather morphogenesis: involvement of anterior-posterior asymmetry and proximal-distal elongation demonstrated with an in vitro reconstitution model. *Development*, 126(12), 2577-2587.

Wong, R. Y., Lamm, M. S., & Godwin, J. (2015). Characterizing the neurotranscriptomic states in alternative stress coping styles. *BMC Genomics*, 16(1), 425.

Wu, L., Fan, J., & Belasco, J. G. (2006). MicroRNAs direct rapid deadenylation of mRNA. *Proceedings of the National Academy of Sciences of the United States of America*, 103(11), 4034-4039.

Wu, P., Jiang, T. X., Suksaweang, S., Widelitz, R. B., & Chuong, C. M. (2004). Molecular shaping of the beak. *Science*, 305(5689), 1465-1466.

Xie, C., Mao, X., Huang, J., Ding, Y., Wu, J., Dong, S., ... & Wei, L. (2011). KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Research*, 39(suppl_2), W316-W322.

Xue, Z., Huang, K., Cai, C., Cai, L., Jiang, C. Y., Feng, Y., ... & Liu, J. Y. (2013). Genetic programs in human and mouse early embryos revealed by single-cell RNA [thinsp] sequencing. *Nature*, 500(7464), 593-597.

Yu, M., Wu, P., Widelitz, R. B., & Chuong, C. M. (2002). The morphogenesis of feathers. *Nature*, 420(6913), 308-312.

Yu, M., Yue, Z., Wu, P., Wu, D. Y., Mayer, J. A., Medina, M., ... & Chuong, C. M. (2004). The developmental biology of feather follicles. *The International Journal of Developmental Biology*, 48, 181.

Zhang, B., & Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. *Statistical Applications in Genetics and Molecular Biology*, 4(1).

Zhang, G., Li, C., Li, Q., Li, B., Larkin, D. M., Lee, C., ... & Ödeen, A. (2014). Comparative genomics reveals insights into avian genome evolution and adaptation. *Science*, 346(6215), 1311-1320.

Zhang, L., Nie, Q., Su, Y., Xie, X., Luo, W., Jia, X., & Zhang, X. (2013). MicroRNA profile analysis on duck feather follicle and skin with high-throughput sequencing technology. *Gene*, 519(1), 77-81.

Zhou, X., Kao, M. C. J., & Wong, W. H. (2002). Transitive functional annotation by shortest-path analysis of gene expression data. *Proceedings of the National Academy of Sciences*, 99(20), 12783-12788.

Zimek, A., & Weber, K. (2005). Terrestrial vertebrates have two keratin gene clusters; striking differences in teleost fish. *European Journal of Cell Biology*, 84(6), 623-635.

APPENDIX A

CAPTIONS OF ADDITIONAL FILES

Additional File 2.1. Is reprinted with permission from Zhang et al (2014) and lists the copy numbers of the Alpha (α)- and beta (β)-keratin for each species of mammal, reptile and bird used in this study. Copy numbers for mammalian α-keratins (Vandebergh and Bossuyt, 2012), green anole β-keratins (Dalla Valle et al., 2010), green sea turtle β-keratins (Li et al., 2013) and crocodilian β-keratins (Greenwold and Sawyer, 2013) were obtained from their respective studies. Species in bold have an aquatic or semi-aquatic lifestyle and those with an asterisk next to their name have a predatory lifestyle.

Additional File 2.2. Contains tables listing the number of α- and β-keratins determined by our annotation and phylogenies for each species in this study.

Additional File 2.3. Contains the 48 bird genome scale phylogeny and is reprinted with permission from Jarvis et al. (2014).

Additional File 2.4. Is reprinted and modified with permission from Zhang et al. (2014) and illustrates the molecular phylogeny of β-keratins. This figure displays the maximum likelihood phylogeny of β-keratins from the green anole lizard, green sea turtle, American alligator and the 48 birds. The green anole lizard β-keratins formed a significant clade and were subsequently used as an outgroup. Colored clades are statistically significant except for the scale β-keratins, which only have a subset forming a significant clade. The green sea turtle β-keratin genes are found in the keratinocyte and claw β-keratin clades while the American alligator β-keratin genes are found in the keratinocyte, scale and claw β-keratin clades. The feather β-keratin clade is composed of only avian β-keratins. The feather β-keratin clade is further annotated based on genomic loci (chromosomes) of the chicken and/or zebra finch feather β-keratins. Chr. is an abbreviation for chromosome and F-L indicates feather-like β-keratins. Chr. 27 A corresponds to genes found on the 5'

most array of feather β-keratins on zebra finch microchromosome 27 (see Figure 3 of Greenwold and Sawyer [2010]) and related genes, while Chr. 27 B are the genes on the 3' most array of feather β-keratins on zebra finch microchromosome 27 and chicken chromosome 5 and microchromosome 27.

Additional File 2.5. Contains tables listing the differentially expressed genes and their p-value, fold change and regulation for each sample comparison.

Additional File 2.6. Contains tables listing species and identification number for the sequences used as queries in genome searches.

Additional File 3.1. Bootstrap hierarchical clustering of the 31 individual microarray samples. Approximately Unbiased (AU) p-value and Bootstrap Probability (BP) are listed on the branches.

Additional File 3.2. This file lists the genome search results of the 737 chicken miRNA genes and the 226 differentially expressed miRNA genes for the 60 tetrapods. The numbers listed are the number of BLAST hits. Highlighted cells indicate no BLAST hits for the respective species and miRNA gene. The last sheet lists the scientific Latin species name and the corresponding English name and 5 letter abbreviation for each species.

Additional File 3.3. The “Main Table” sheet lists the microarray comparison group in which the 226 differentially expressed chicken miRNA genes are differentially expressed. The “pre and mature miRNA” sheet lists the differentially expressed pre-miRNAs and their correlated mature miRNAs. The remaining sheets list the p-value, fold change value and direction of regulation for the differentially expressed pre-miRNAs in the 14 comparison groups.

Additional File 3.4. Venn diagrams of the overlap of differentially expressed miRNAs between comparison groups. The top left figure illustrates the overlap miRNAs for intra-tissue comparison groups between day 17 and day 19, top right illustrates inter-tissue comparison groups at day 17 and bottom figure illustrates the inter-tissue comparison groups at day 19.

Additional File 3.5. This file lists the RNAhybrid results of the 226 differentially expressed miRNA genes sorted by microarray comparison group. The mRNA genes are listed as NCBI GenBank accession numbers. The values in parentheses represent the corresponding miRNA-mRNA duplex minimum free energy (mfe) value and are sorted in descending order. The DF19 and WF19 comparison group is not included due to the lack of differentially expressed miRNAs in that comparison group.

Additional File 3.6. The "Summary" sheet lists the BLAST hit numbers of the predicted target site for the non-keratin target genes listed in Table 3.1. The remaining worksheets list the detailed BLAST hit information including the cutoff for each of the genes in Table 3.1.

Additional File 4.1. Illustrates the sample clustering based on individual sample expression profiles. Each array sample was named as the tissue type with the days of embryonic stages. The y-axis (Height) is the distance metric calculated by average linkage hierarchical clustering method. There are no outliers and most of the samples with the same tissue type and embryonic stage cluster together.

Additional File 4.2. We apply a binary system (1 for positive and 0 for negative) to define and quantify the qualitative traits to measure the external trait based on the

microarray expression profiles. This is the spreadsheet of this particular trait-defining strategy for 31 samples.

Additional File 4.3. We performed the analysis of network topology to facilitate choosing soft threshold power β for constructing the WGCNA network. The left panel demonstrates the scale-free topology fit index (vertical axis) as the soft threshold power β (horizontal axis) varies. The right panel displays the mean connectivity (degree, vertical axis) as the soft threshold power β (horizontal axis) varies. The soft threshold power was selected based on the criterion of approximate scale-free topology which has high scale-free topology fit index (normally above 0.8) and the saturation of the mean connectivity was reached by the lowest power β.

Additional File 4.4. For the scale and feather traits, we list all modules having moderate Module-Trait correlations ($r > 0.5$, $p < 0.01$) individually, we also highlighted the modules also having moderate MM-GS correlations ($r > 0.5$, $p < 0.01$).

Additional File 4.5. To incorporate the external traits, gene significance (GS) of the genes having Entrez Gene ID annotations for Feather ("Feather trim" worksheet) and Scale (Scale trim worksheet) traits were sorted based on the absolute GS value. We also list the corresponding module color and the p-value of GS.

Additional File 4.6. We performed the KEGG pathway and Gene Ontology enrichment analysis individually for 11 modules having both correlated Module-Trait and MM-GS relationships (red color modules in Supplemental Table 2). The 2 worksheets listed all significantly enriched KEGG pathways and GO annotations with details based on the corrected p-value

Additional File 4.7. Demonstrates all epidermal development related genes (EDRGs) in brown and black modules. We also list the number of the edges (connectivity) for these genes. The rank of these EDRGs based on the connectivity number among all genes having gene symbol annotation and the top percentage were listed

APPENDIX B

COPYRIGHT PERMISSION FOR CHAPTER 2

Dear Dr. Bao,

Thank you for contacting BioMed Central.

The open access articles published in BioMed Central's journals are made available under the Creative Commons Attribution (CC-BY) license, which means they are accessible online without any restrictions and can be re-used in any way, subject only to proper attribution (which, in an academic context, usually means citation).

The re-use rights enshrined in our license agreement (http://www.biomedcentral.com/about/policies/license-agreement) include the right for anyone to produce printed copies themselves, without formal permission or payment of permission fees. As a courtesy, however, anyone wishing to reproduce large quantities of an open access article (250+) should inform the copyright holder and we suggest a contribution in support of open access publication (see suggested contributions at http://www.biomedcentral.com/about/policies/reprints-and-permissions/suggested-contributions).

Please note that the following journals have published a small number of articles that, while freely accessible, are not open access as outlined above: Alzheimer's Research & Therapy, Arthritis Research & Therapy, Breast Cancer Research, Critical Care, Genome Biology, Genome Medicine, Stem Cell Research & Therapy.

You will be able to find details about these articles at http://www.biomedcentral.com/about/policies/reprints-and-permissions

If you have any questions, please do not hesitate to contact me.

With kind regards,

Joel Lagmay

Global Open Research Support Executive

Global Open Research Support

Springer Nature

T +44 (0)203 192 2009

www.springernature.com

APPENDIX C

COPYRIGHT PERMISSION FOR CHAPTER 3

ELSEVIER LICENSE TERMS AND CONDITIONS

This Agreement between University of South Carolina -- Weier Bao ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License Number: 4153421196202

License Date: Jul 20, 2017

Licensed Content Publisher: Elsevier

Licensed Content Publication: Gene

Licensed Content Title Expressed miRNAs target feather related mRNAs involved in cell signaling, cell adhesion and structure during chicken epidermal development

Licensed Content Author: Weier Bao, Matthew J. Greenwold, Roger H. Sawyer

Licensed Content Date: Oct 15, 2016

Licensed Content Volume: 591

Licensed Content Issue: 2

Licensed Content Pages: 10

Start Page: 393

End Page: 402

Type of Use: reuse in a thesis/dissertation

Intended publisher of new work: other

Portion: full article

Format: both print and electronic

Are you the author of this Elsevier article? Yes

Will you be translating? No

Title of your thesis/dissertation: MOLECULAR EVOLUTION OF MAJOR EPIDERMAL STRUCTURE GENES AND AN INTEGRATIVE TRANSCRIPTOME ANALYSIS OF CHICKEN EPIDERMAL EMBRYOGENESIS

Expected completion date: Aug 2017

Estimated size (number of pages): 120

Requestor Location: University of South Carolina

Attn: University of South Carolina

Publisher Tax ID: 98-0397604

Total: 0.00 USD

APPENDIX D

COPYRIGHT PERMISSION FOR CHAPTER 4

SPRINGER LICENSE TERMS AND CONDITIONS

This Agreement between University of South Carolina -- Weier Bao ("You") and Springer ("Springer") consists of your license details and the terms and conditions provided by Springer and Copyright Clearance Center.

License Number: 4153420792309

License date: Jul 20, 2017

Licensed Content Publisher: Springer

Licensed Content Publication: Functional & Integrative Genomics

Licensed Content Title: Using scale and feather traits for module construction provides a functional approach to chicken epidermal development

Licensed Content Author: Weier Bao

Licensed Content Date: Jan 1, 2017

Type of Use: Thesis/Dissertation

Portion: Full text

Number of copies: 1

Author of this Springer article: Yes and you are a contributor of the new work

Title of your thesis / dissertation: MOLECULAR EVOLUTION OF MAJOR EPIDERMAL STRUCTURE GENES AND AN INTEGRATIVE TRANSCRIPTOME ANALYSIS OF CHICKEN EPIDERMAL EMBRYOGENESIS

Expected completion date: Aug 2017

Estimated size (pages): 120

Requestor Location: University of South Carolina

United States

Attn: University of South Carolina

Billing Type: Invoice

Total: 0.00 USD